

# Comparison of modern open-source Visual SLAM approaches

Dinar Sharafutdinov · Mark Griguletskii · Pavel Kopanев · Mikhail Kurenkov ·  
 Gonzalo Ferrer · Aleksey Burkov · Aleksei Gonnochenko · Dzmitry Tsetserukou

Received: date / Accepted: date

**Abstract** Simultaneous localization and mapping (SLAM) is one of the fundamental areas of research in robotics and environment reconstruction. State-of-the-art solutions have advanced significantly in terms of mapping quality, localization accuracy and robustness. It becomes possible due to modern stable solvers in the back-end, efficient outlier rejection techniques and diversified front-end: unique features, topologically segmented landmarks, and high-quality sensors. Among the variety of open-source solutions, several promising approaches provide results which are difficult to be reproduced on standard datasets, especially if there is no description for dataset adaptation. The goal of the article is to figure out, which techniques of robots' localization are the most promising for further use in related disciplines for engineers and researchers. The main contribution is a comparative analysis of state-of-the-art open-source Visual SLAM methods in terms of localization precision for versatile environments. The algorithms are assessed based on accuracy, computational performance, robustness and fault tolerance. Additionally, the survey and comparison of the datasets used for methods evaluation are provided as well as practical recommendations of usage scenarios for further research.

---

Skolkovo Institute of Science and Technology:  
 Dinar Sharafutdinov E-mail: dinar.sharafutdinov@skoltech.ru, Mark Griguletskii E-mail: mark.griguletskii@skoltech.ru, Pavel Kopanev E-mail: pavel.kopanev@skoltech.ru · Mikhail Kurenkov E-mail: mikhael.kurenkov@skoltech.ru, Gonzalo Ferrer E-mail: G.Ferrer@skoltech.ru, Dzmitry Tsetserukou E-mail: D.Tsetserukou@skoltech.ru  
 Sberbank Robotics Lab:  
 Aleksey Burkov E-mail: burkov.a.m@sberbank.ru, Aleksei Gonnochenko E-mail: gonnochenko.a.s@sberbank.ru

**Keywords** SLAM · VIO · Benchmarking · SLAM comparison

## 1 Introduction

Simultaneous localization and mapping (SLAM) is an important task in computer robotics, computer vision and environment reconstruction. Robots need to estimate their current position and surrounding map during operation. SLAM algorithms require measurements between sequential positions of the robot (aka odometry) and between the robot and landmarks. Thus, the goal of SLAM is to reconstruct a consistent map of the environment and localize a robot on it. The problem becomes much more difficult in an unknown environment with no prior information.

There are many open-source solutions that have been developed over the last 20 years. However, it is a complex problem of selecting an open-source SLAM system among the options available. The algorithms are implemented differently and tested on diverse datasets. At the same time, a good common evaluation of SLAM systems is essential for robotics engineers. They must know weaknesses and strengths to properly use SLAM solutions in robotic systems.

There are several approaches to solve the visual SLAM problem. Main methods are feature-based [49] and direct methods [24]. Feature-based methods use key points. In these methods, key points are saved to local submaps or key-frames and then a graph optimization approach is applied. Whereas direct-SLAM methods save all pixels of images on local maps and use photometric losses for measurement error estimations. Modern SLAM approaches are very different and usually highly modular. They consist of front-end and back-

end. Moreover, some state-of-the-art methods solve the SLAM problem with dynamic objects.

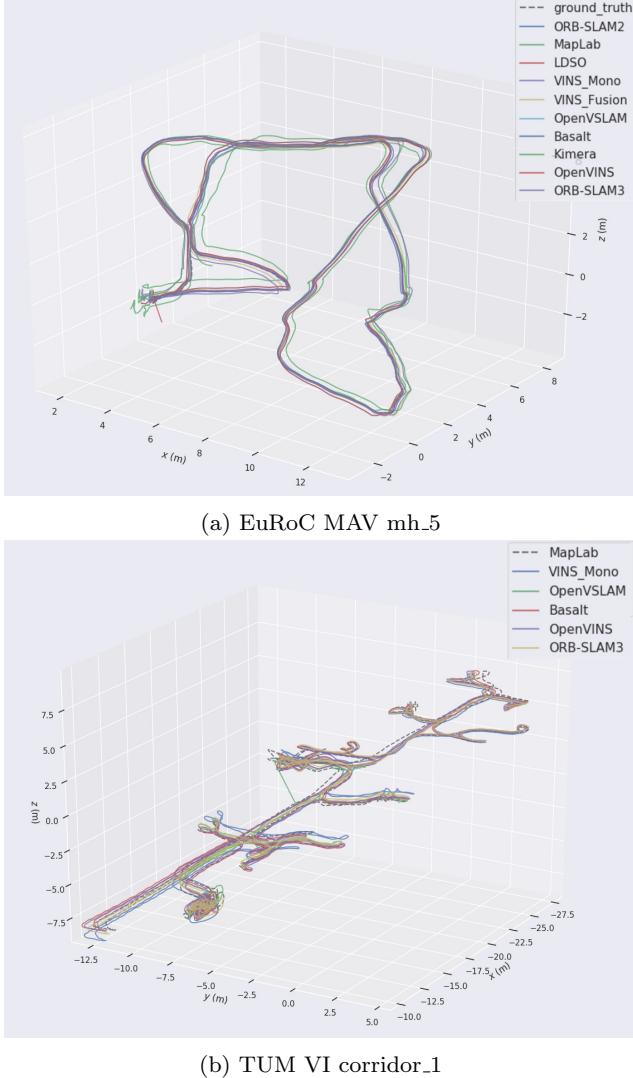


Fig. 1: Examples of resulting trajectories.

During the last years, SLAM systems for different sensors apart from RGB cameras have been developed. For example, RGB-D (depth) and stereo cameras [50] as well as IMU measurements [56] or transformations between lidar point clouds [6]. The combination of different information about the environment leads to a consistent solution to the SLAM problem. Thus, the approaches must be implemented and evaluated on the datasets with versatile sensors. This paper focuses on feature-based solutions: ORB-SLAM 2/3, MapLab, LDSO, VINS-Mono, VINS-Fusion, Open VSLAM, Basalt, Kimera, Open VINS, and DRE. Section 3.1 (Choice of algorithms) describes the choice of algorithms in detail. A full comparison of localization acc-

curacy of different algorithms as well as memory and computational resources is represented in section 4 (Results). The practical aspects, description of experiments and datasets review could be found in section 3 (Experimental setup). The facilities of original algorithms such as availability of documentation, examples on popular datasets, the convenience of the interface, ability to change the parameters of algorithms and presence of Docker/ROS wrappers (table 1 for more details) are qualitatively evaluated.

Table 1: Practical facilities of algorithms.

Framework	Documentation	Support	Usage examples on popular datasets	Docker	ROS
ORB-SLAM2	github	-	+	-	+
MapLab	github	-	+	-	+
LDSO	github	+	+	-	-
VINS-Mono	github	-	+	+	+
VINS-Fusion	github	-	+	+	+
OpenVSLAM	web-page, github	+	+	+	+
Basalt	github	+	+	+	-
Kimera	github	+	EuRoC only	+	+
OpenVINS	web-page, github	+	+	-	+
ORB-SLAM3	github	+	+	-	+
DRE	github	-	-	-	+

Modern SLAM technologies achieve significant results [20] but several challenges exist [14]. The main problems are long-term stability, scalability, sensor biases and miscalibrations, computation and memory requirements, dynamic objects, ambiguous scenes, and accuracy requirements. Since one of the most popular areas of SLAM usage is mobile robotics in the outdoor environment, critical issues are dynamic objects and map size. Thus, SLAM solutions should take these challenges into account and modern datasets and benchmarks need to be focused on these problems.

Nowadays, technical research and papers suffer a common issue with the reproducibility of their methods and results. This is called the “reproducibility crisis”. Frequently, it is very complicated to repeat the same results as described in paper or documentation. But this might be critical for researchers and engineers. Fast reproducibility allows testing and evaluating modern SLAM methods in their environments and choosing the appropriate setup for a particular task.

There is a number of datasets and benchmarks for SLAM evaluation. Some of them are well described in a relevant work [46]. A dataset should reflect SLAM challenges, for example, dynamic objects, ambiguous scenes, and huge spaces.

The purpose of the current paper is to give a general introduction to algorithmic state-of-the-art solutions for visual SLAM methods and visual/visual-inertial odometry, to evaluate these algorithms on the same platform, compare the approaches on open datasets and give practical advice for SLAM researchers. Figure 1 il-

lustrates the trajectories for visualization purposes and alignment consistency.

The contribution of the article is a comparative analysis of modern Visual SLAM solutions which consists of results reproduction and multilateral comparison on various open datasets by evaluating localization accuracy, reproducibility, usability, needed computational resources and robustness to various scenarios. These aspects allow us to analyze the advantages and disadvantages of methods in different conditions and understand which of them are suitable for various situations. Additionally, the article provides a practical comparison and description of datasets used for methods evaluation as well as hints for researchers for choosing Visual SLAM methods supported by published GitHub repository<sup>1</sup> with Dockers of the algorithms.

## 2 Related work

For localization and mapping, many sensors are used as input data. For example, 3D LIDARs, 2D LIDARs, depth cameras, stereo cameras, global and rolling shutters cameras. LIDARs are quite precise but very expensive. Moreover, it needs to emit light signals to estimate distance. Therefore, LIDARs are active sensors. One of the most beneficial sensors for robots is the rolling shutter camera. It is cheap, dense and passive. They do not need to emit light and only receive surrounding information. Global shutter cameras might be more convenient, but the price is higher. Stereo and depth cameras such as D435 are also appropriate and very popular. A comprehensive comparison of modern sensors for this problem was made by R. Singh and K. S. Nagla in [66]. All of them can be used in SLAM pipelines, but they have different features which are important while solving the SLAM problem 1. The last is formulated as a maximization of conditional probability ( $P$ ) of the robot's states ( $X$ ) and landmarks ( $M$ ) given sensor measurements ( $Z$ ) and information matrices ( $\Omega$ ) which is equivalent to minimization of minus logarithm of  $P$  in case of Gaussian structure of sensors' noises. It is equivalent to solving a non-linear least-squares problem or kernel-based ( $\rho$ ) minimization for robustness in case of outliers in measurements.

$$\begin{aligned} \hat{X}, \hat{M} &= \arg \max_{X, M} P(X, M|Z) \\ &\simeq \arg \min_{X, M} \{-\log(P(X, M|Z))\} \\ &\equiv \arg \min_{X, M} \sum_{ij} \rho(X_{ij}, M_{ij}, \Omega_{ij}) \end{aligned} \quad (1)$$

<sup>1</sup> SLAM-Dockers <https://github.com/KopanevPavel/SLAM-Dockers>

Worth mentioning, that the main difference between SLAM and Visual/Visual-inertial odometry is that the last only updates the robot's states ( $X$ ) whereas pure SLAM additionally updates the map itself ( $M$ ). This paper evaluates accuracy in terms of robot localization but not map quality. Apart from visual sensors, inertial measurements are also being used as input: wheel odometry and IMU measurements are actively utilized in robotics. SLAM has been extensively studied during the last decades. For example, Cadena et al. [14] presented a good overview of modern SLAM technologies and the challenges which SLAM methods. According to another review [15], today SLAM is going into the spatial artificial intelligence age. This means that more deep learning techniques are used as a solution.

### 2.1 Sparse visual SLAM

The history of feature-based SLAM or sparse visual SLAM began with the history of visual SLAM. One of the first successful approaches which solved the problem of simultaneous localization and mapping is a method presented by Davison et al. [18]. In this work, the MonoSLAM algorithm had been introduced. It was able to find the camera position from the input images. The authors managed to speed up the solution of the SfM (Structure from Motion) problem by using methods based on probabilistic filters. The algorithm can obtain the position of point landmarks and the trajectory of the camera in real time.

There are several disadvantages of the MonoSLAM method. The first problem is the filtering approach which badly scales. The other disadvantage of the MonoSLAM algorithm was the necessity to update the trajectory and the feature positions sequentially in the same thread. To solve this problem, Klein et al. [43] proposed the PTAM (Parallel Tracking and Mapping) algorithm which achieves real-time performance on an algorithm that executes full batch updates. Additionally, authors of the work [67] showed that the solutions based on bundle adjustment [43] are superior to filter-based ones.

The key contribution in the field of solving the problem of visual SLAM was made by the ORB-SLAM algorithm which was described in this article [49]. The authors of the paper integrated all previous methods together in the same frame and combined the key elements of visual SLAM methods: Bags of Binary Words [30] visual terrain recognition algorithm, ORB [60] algorithm for quick detection and description of key points and g2o [45] optimizer. Figure 2 shows general scheme of the method. There are also other backend optimizers that researchers highly used. For example, GTSAM [19]

and Ceres [1]. However, this article focuses more on different frontend parts of SLAM systems.

ORB-SLAM has three parallel processes. The first process is the construction of the local camera trajectory by matching the observed key points to the local map. The second process builds a local map and solves the local bundle adjustment problem. And the last process finds loop closures. Moreover, it can trigger a full optimization of the entire camera trajectory. The authors of ORB-SLAM also proposed a second version of their algorithm: ORB-SLAM2 [50]. This algorithm allows using stereo and depth cameras.

ORB-SLAM method suffers from drifts of a scale. In order to reveal the geometric scale another source of information is often used, for example, an IMU sensor. The article [57] introduces the VINS algorithm which uses the Kalman filter to merge sequential images and IMU data to calculate odometry. As a logical continuation of [55], the authors started to use global optimization to relocate the camera.

The development of ORB-SLAM2 can be considered as SOFT-SLAM [17], the authors of this algorithm made the following changes:

- Instead of the more computationally complex bundle adjustment for localization, SOFT visual odometry is used, which achieves an error of about 0.8% relative to the distance travelled.
- The streams responsible for mapping and odometry are separated from each other, thus the visual odometry stream is not blocked by the thread responsible for mapping, which leads to a more stable processing time for incoming frames. In addition, unlike ORB-SLAM2, the algorithm is completely deterministic, i.e. always returns the same result for the same input.
- SOFT key points are used for loop closure (similar to ORB-SLAM2), which makes the system highly efficient, simple and reliable while achieving sub-pixel precision. Despite the fact that SOFT key points are not invariant with respect to rotation, the authors show on public datasets that in practice, loop closure occurs often enough that this drawback has little effect on the final result.

It should also be noted that there is ORB-SLAM2-CNN [71], which adds semantic information received from the neural network to the ORB-SLAM2 framework.

Another key point method worth mentioning is VINS-Mono [58]. This method is focused on processing monocular data augmented with data from IMU.

The article [10] questions the visual SLAM approach based on global optimization of the entire trajectory

and key point positions, that is, based on bundle adjustment. As an alternative, the authors of the article propose the L-infinity SLAM method. The essence of this method is that the robot first separately finds the camera rotations by averaging the relative rotations, after which it solves the optimization problem to find the three-dimensional position of the cameras and key points with known camera orientations. One can also use the rotation averaging described in the publication [10] for direct methods, and as additional restrictions for other ways to solve the visual SLAM problem.

## 2.2 Dense visual SLAM

Unlike methods for solving the problem of visual SLAM, which are based on key points, dense or direct methods do not use algorithmic features in their work. They use photometric error minimization for each pair of images to find the orientation and position of the camera in 3D space.

The first work that suggested using direct methods was DTAM [52]. For each image, the DTAM algorithm generated a dense 3D map for each pixel. The next algorithm was KineticFusion [51], which uses a depth camera to build a dense 3D map. The authors use TSDF (truncated signed distance function) to describe each pixel, and ICP (iterative closest point) algorithm to map each depth image to a map.

The next milestone in the development of direct methods was the ElasticFusion [74] algorithm. This algorithm uses a direct representation of the surfaces of RGB-D cameras. For data fusion, this algorithm uses a non-rigid deformation model. Also in this algorithm, there is no global optimization of the graph of camera positions, and most of the steps take place on a graphics accelerator (GPU).

In parallel with the development of KineticFusion and ElasticFusion, two other teams were developing semi-dense visual odometry (SVO) [25, 28]. This algorithm can run in real-time on a robot processor (CPU). The SVO algorithm does not use all the pixels in the image, but only those that have a negligible gradient. The development of ideas from the publication [25] leads to the creation of a key work in this area - LSD-SLAM [24]. LSD-SLAM uses  $\text{sim}(3)$  - a metric to solve a problem with an uncertain scale, and also uses probabilistic inference to determine the error in constructing three-dimensional maps. The LSD-SLAM algorithm consists of three steps:

1. Get the image and determine the local offset relative to the current keyframe.

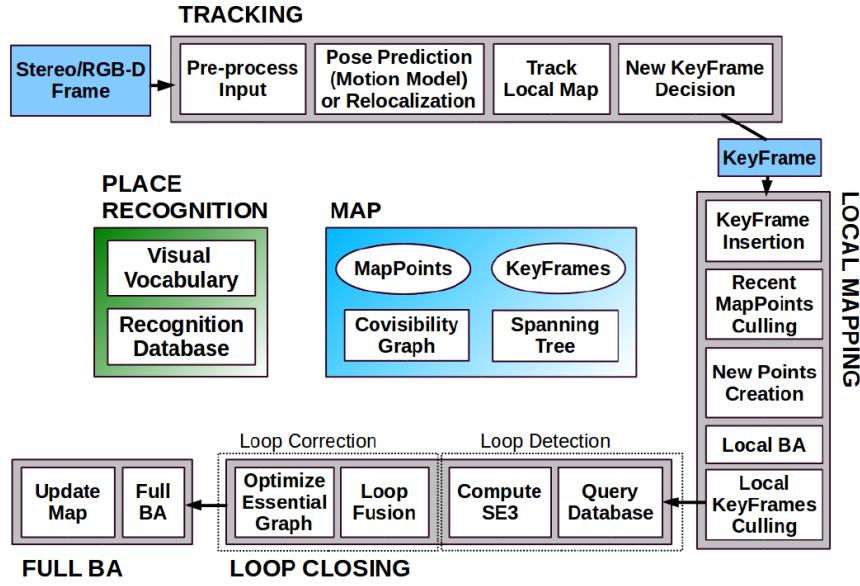


Fig. 2: The scheme of the ORB-SLAM algorithm [49].

2. Update or create a new keyframe. If the current keyframe is updated, probabilistically merge the stereo depth data into the current keyframe. In case of creating a new keyframe, propagate the previous depth map to the new keyframe.
3. Updating the global map using position graph optimization. The edges in the graph are searched using the sim (3) metric, and the optimization is performed using the g2o library.

The LSD-SLAM development team continued the development of direct methods and proposed the DSO (direct sparse odometry) [22] method. Their method optimizes not only the photometric error but also the geometric error simultaneously. Also, instead of assuming smoothness, the authors use probabilistic pixel sampling. In LDSO [31], they added graph optimization to get a complete solution to the visual SLAM problem. In fact, LDSO is a combination of the ORB-SLAM and LSD-SLAM approaches.

Separately, it should be noted that the authors of DSO in 2019 published a modification of DSO [63], adapted for processing frames obtained using a camera with a floating shutter. In addition to the photometric bundle adjustment, IMU data are also used as additional constraints for the optimization problem. Due to the fact that there is no selection of key points, this method (like DSO) can work not only with pixels representing edges and corners but with any pixels that have a sufficient gradient.

Many improvements to direct methods have also been proposed recently. For example, CodeSLAM [8]

suggests using autoencoders to train a more compact map representation for direct methods. Also, in the work KO-fusion [37] it was proposed to use odometry and kinematics data obtained directly from the robot arm.

### 2.3 Dynamic visual SLAM

Usually, SLAM solutions suppose that scene is almost static or with a low level of dynamics. The simplest way of discarding outliers is to use RANSAC [26]. In reality, almost in every scenario, there will be lots of dynamic objects. Both indoor and outdoor scenes are full of people, animals, cars, pedestrians, bicycles and other dynamic objects. Therefore, it leads to changes in feature maps. This is especially important for loop closure detection. Additionally, in the case of a high level of dynamics, most slam solutions may show inconsistent results because of a lack of reliable features. There are several problems with dynamic objects. Firstly, the algorithms should detect the dynamic object and not use it for the trajectory estimation as well as for the mapping process. A possible solution here is to delete these objects from the image or replace them with a background. Popular tools for that are multi-view geometry [76], semantic segmentation or detection neural networks [5], scene [4] or optical flow and detection of background or foreground [75]. Secondly, there is a need of building a consistent map with or without dynamic objects. Solving the problem of lack of data due to the deletion of dynamic objects is an important as-

pect. Additional sensors, objects completion and reconstruction in combination with different map construction algorithms might help [21] [61]. Dynamic SLAM algorithms often represent complex systems consisting of many parts and working not in real-time.

### 3 Experimental setup

There were several main goals in the experimental stage. First of all, the goal was to test the algorithms which are available to everyone. Only those SLAM solutions that have a link to the page with code and instructions were used. Secondly, it was decided to take one or few algorithms from each group of SLAM methods, instead of doing experiments with all the possible solutions. It is crucial to see the results of algorithms that work with monocular, stereo and RGB-D cameras. Another essential part is understanding how additional data such as IMU or wheel encoder improves the results. The last point here is to try the solutions based on deep learning methods and with various front-end parts. Thirdly, there is the main use case that suits broad possible robotics tasks. Wheel robot with cameras and other sensors moves in an inside/outside environment. It moves among dynamic objects and features-less areas. The robot also visits the same places several times (loop closures), and the traversal could be relatively long (30 minutes). These factors affect the datasets that were used for tests.

#### 3.1 Choice of algorithms

##### 3.1.1 Sparse and dense SLAM

ORB-SLAM2 [50] was used as a baseline sparse SLAM solution. It is quite popular among researchers, and it proved efficiency in many scenarios. The second version is an expansion of the first one, hence ORB-SLAM1 [49] has not been used. Peculiarities of the algorithm were covered in the literature review section.

Another algorithm here is an OpenVSLAM [70]. Generally, it is still ORB-SLAM but with additional features from ProSLAM and UcoSLAM. The main value of this solution is practical extensions such as support of the various type of camera models (perspective, fisheye, equirectangular), map saving and creation features, and finally convenient documentation and UI.

The most modern approach which has been considered is the recently published ORB-SLAM3 [11]. This sparse SLAM system is a natural development of the

previous version of ORB-SLAM. Now it is a visual-inertial approach with an updated place recognition module (high-recall place recognition with geometrical and local consistency check) as well as with the support of pinhole and fish-eye camera models. Moreover, it allows localization in a multi-map setup.

Also, one of the dense SLAM methods have been chosen for comparison. LDSO [31] is a graph-based dense SLAM system with loop closure detection. It is based on Direct Sparse Odometry (DSO) [23]. LDSO is robust and suitable for feature-poor environments utilizing any image pixel with sufficient intensity gradient. It uses a feature-based bag-of-words for loop closure detection and was validated on several open-source datasets.

##### 3.1.2 Visual-Inertial Odometry and SLAM

A typical Visual-Inertial Odometry system implies the usage of a camera and IMU sensors. OpenVINS [34] is Multi-State Constraint Kalman Filter (MSCKF) based VIO estimator. It uses IMU in the propagation step of the filter and camera data in the update step. In the current implementation OpenVINS also has a secondary loosely coupled loop closure thread based on VINS-Fusion [55]. Additionally, OpenVINS has an interface wrapper for exporting visual-inertial runs into the ViMap structure taken by maplab [62].

VINS-Mono [56] and VINS-Fusion is graph-based VIO approaches. VINS-Fusion is an extension of VINS-Mono and supports multiple visual-inertial sensor types (mono camera + IMU, stereo cameras + IMU, even stereo cameras only). Also, it supports global sensors like GPS and Barometer and has a global graph optimization module. Additionally, VINS-Fusion supports loop closure.

Basalt [73] is other graph-based VIO approach. It has a lot of similarities with previously described VINS-Fusion (KLT feature tracking and Gauss-Newton non-linear optimization). But regarding map optimization, this approach utilizes ORB features.

Kimera [3] is an open-source C ++ library that implements one way of solving SLAM problem in real-time. The modular structure includes: a VIO module with a GTSAM-based [19] VIO approach, using IMU-preintegration [27] and structureless vision factors [13] for fast and accurate state assessment, a robust position graph optimizer (RPGO) for global trajectory estimation which adds a robustness layer that avoids SLAM failures due to perceptual aliasing, and relieves the user from time-consuming parameter tuning, a lightweight 3D mesh module (Kimera-Mesher) for fast 3D mesh reconstruction [59] and obstacle avoid-

ance and a dense 3D semantic reconstruction module (Kimera-Semantics) that builds a more-accurate global 3D mesh using a volumetric approach [54], and semantically annotates the 3D mesh using 2D pixel-wise semantic segmentation based on deep learning methods.

### 3.1.3 Dynamic SLAM

In this section three algorithms have been considered: DynaSLAM [5], DynSLAM [4] and DRE-SLAM [75]. All of them have theoretical potential and relatively popular implementations. DynaSLAM is a successor of ORB-SLAM2 with added front-end part that allows working in a dynamic environment. The solution consists of the segmentation of dynamic objects using CNN in monocular and stereo cases, and a combination of deep neural methods and multi-view geometry in the RGBD case. The important feature is that with DynaSLAM, it is possible to detect a priori dynamic objects, objects that may be static at the moment but the dynamic in essence. Pixel-wise semantic segmentation of potentially movable objects is possible with the help of Mask R-CNN. Then combining segmentation with multi-view geometry allows finding even static objects moved by dynamic ones (e.g. book in the hands of a person). After the deletion of dynamic objects, there are empty regions on images. To solve that, the authors use background inpainting by using information from previous views.

DynSLAM is a more complex solution that solves even more tasks for large-scale dynamic environments. Taking stereo images as the input, they compute depth maps (using ELAS or DispNet) and sparse scene flow (libviso2). Multi-task Network Cascades allow finding dynamic objects. Using obtained information, they do 3D object tracking and reconstruction. The final result is a static map without dynamic objects with the help of InfiniTAM. The last one, DRE-SLAM, is a bit different from the previous ones. Apart from image data, it also uses two-wheel encoders to handle the lack of features in dynamic scenes. It extracts ORB features from RGB-D images, uses YOLOv3 for dynamic object detection and then applies multiview constraint-based pixel culling. Loop closure detection is implemented through a bag-of-words approach. The final map is OctoMap which is constructed by fusing sub-OctoMaps.

Although DynaSLAM and DynSLAM have developed and published solutions, the launch of the algorithms has not been successful. GitHub pages full of issues and implementations are not supported by authors, which makes solving varying problems very difficult. On the other hand, DRE-SLAM ran smoothly

on ROS but it has examples only for data collected by authors. Comparison with other algorithms on popular datasets was difficult because of the lack of wheel encoder data in many of them. Nevertheless, the comparison has been performed with several other algorithms on the DRE dataset.

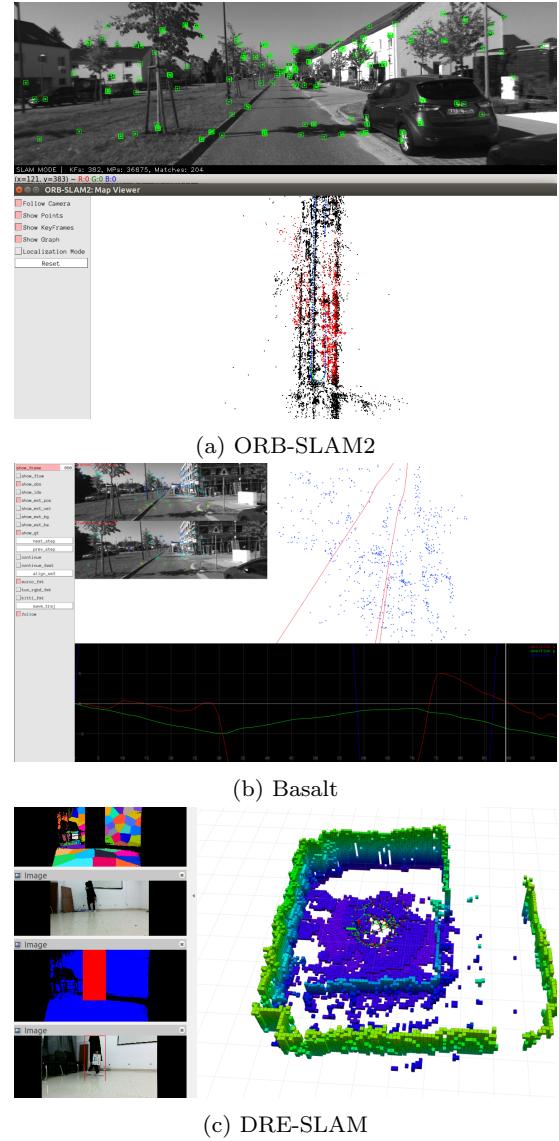


Fig. 3: GUI examples.

## 3.2 A practical overview of open source solutions

During the process of launching and working with algorithms, a number of aspects has been identified that must be taken into account when dealing with SLAM solutions.

First, all algorithms were run on Linux-based systems (most often Ubuntu). Launching on Windows seems to be quite time-consuming and, in general, all examples and instructions from the authors are made for Ubuntu or macOS. Moreover, it was found that a number of algorithms require older versions of Ubuntu (16.04 or 18.04). In some cases, this can be critical for a successful launch.

Secondly, algorithms have different interfaces. The most unified way is to use the Robot Operating System (ROS). In this case, everything depends on the support of the developers of the algorithm. The presence of a separate community, pipeline for data usage, and many examples make ROS a convenient option for launching. Other options are graphical interfaces used by developers (e.g. Pangolin viewer). The quality of the viewer implementation and interface depends on the developer and may differ from one algorithm to another.

Thirdly, the points mentioned above make it especially important to have support from the authors of the SLAM systems. It can be in form of answers to issues on GitHub, availability of documentation and examples of launching. In some cases, the authors have their own website with documentation, or even a channel on Slack to discuss the algorithm and emerging problems [70].

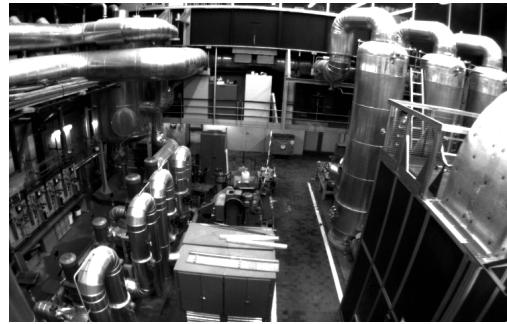
Fourth, the most convenient way to run SLAM algorithms is to use a ready-made Docker image. Just a few authors provide such a solution. To simplify the launch of algorithms and to the benefit of the community, a publicly available [44] repository with docker images has been created and uploaded to the Docker Hub.

### 3.3 Datasets description

Datasets are actively used for SLAM algorithms validation and comparison. They should include ground truth information and have a redundant amount of sensors in order to test different SLAM approaches based on different sensor stacks.

This paper reviews visual and visual-inertial SLAM approaches, and therefore one of the main dataset choice factors was camera image presence. Datasets were examined for the existence of loop closures and additional sensors (IMU, GPS, or wheel encoders). This section does a short description of all the datasets that have been considered. The final choice for the experiments is five datasets: EuRoC MAV, TUM VI, KITI, Open Loris, and DRE dataset. Camera data is shown in Fig. 4.

A commonly used option for evaluating visual and visual-inertial SLAM algorithms is the EuRoC MAV dataset [9], but its image resolution and bit depth are



(a) EuRoC MAV



(b) TUM VI



(c) KITI



(d) OpenLoris-Scene



(e) DRE dataset

Fig. 4: Examples of images from the datasets.

not quite state-of-the-art anymore. Also, this dataset is mostly suitable for Micro Air Vehicles and does not include odometry readings suitable for some SLAM approaches for mobile wheeled robots. SLAM article authors usually test and compare their algorithms on the EuRoC dataset primarily but in many cases SLAM algorithms show good performance on this dataset and bad on other datasets. That will be shown later.

TUM VI [64] is another popular dataset for evaluating visual and visual-inertial SLAM algorithms. It includes more varied scenes including indoor and outdoor environments and longer sequences than the EuRoC

Table 2: Sensor setup for dataset collection.

Dataset name	Camera	GPS	IMU	Lidar	Wheel odometry	Additional sensors
EuRoC MAV	Aptina MT9V034 global shutter, WVGA monochrome (2x20 Hz)	x	ADIS16448 (200 Hz)	x	x	x
TUM VI	1 stereo gray (20 Hz)	x	BMI160 (200 Hz)	x	x	x
KITTI	2x Point Grey FL2-1453C-C global shutter (15 Hz)	OXTS RT 3003 (250 Hz)	OXTS RT 3003 (250 Hz)	Velodyne HDL-64E (5-20 Hz)	x	x
OpenLoris-Scene	RealSense D435i (30 Hz) RealSense T265 (30 Hz)	x	RealSense (60-400 Hz)	Hokuyo UTM30LX (40 Hz)	Odometer (20 Hz)	x
DRE dataset	Kinect 2.0 (20 Hz)	x	x	x	Wheel encoders (100 Hz)	x
KAIST	2x FLIR FL3-U3-20E4C-C global shutter (2x10 Hz)	U-Blox EVK-7P (10 Hz)	Xsens MTi-300 (200 Hz)	Velodyne VLP-16 (10 Hz), SICK LMS-511 (100 Hz)	2x RLS LM13 encoders (100 Hz)	VRS-GPS, Three-axis FOG, Altimeter
Malaga Urban	1 stereo RGB (20 Hz)	DELUO (1 Hz)	xSens MTi (100 Hz)	3x Hokuyo UTM-30LX (25 Hz), 2x SICK LMS-200 (75 Hz)	x	x
Oxford RobotCar	BBX3-13S2C-38 global shutter (16 Hz), 3x GS2-FW-1455C-C global shutter (11.1 Hz)	NovAtel SPAN-CPT (50 Hz)	NovAtel SPAN-CPT (50 Hz)	2 x SICK LMS-151 (50 Hz) 1 x SICK LD-MRS (12.5 Hz)	x	x
RPNG OpenVINS	Aptina MT9V034 (60 Hz)	x	ADIS16448 (200 Hz)	x	x	x
Segway DRIVE	RealSense ZR300 (30 Hz)	x	BMI055 (250 Hz)	Hokuyo UTM30LX (40 Hz)	x	x
TUM RGB-D	Kinect (30 Hz)	x	x	x	x	x
UMich NCLT	6 RGB (omni) (5 Hz)	Garmin 18x (5 Hz)	Microstrain 3DM-GX3-45 (100 Hz)	Velodyne HDL-32E (10 Hz) Hokuyo UTM-30LX (40 Hz) Hokuyo URG-04LX (10 Hz)	Wheel encoders	FOG, RTK GPS
Zurich Urban MAV	RGB rolling shutter (30 Hz)	GPS	IMU (10 Hz)	x	x	x

Table 3: Dataset main features.

Dataset name	Year	ROS bag	Ground truth	Indoor / Outdoor
EuRoC MAV	2016	+	Motion capture system (acc. $\approx$ 1 mm)	Indoor
TUM VI	2018	+	Motion capture system (acc. $\approx$ 1 mm)	Indoor / Outdoor
KITTI	2013	+ (parser is provided)	OXTS RT 3003 inertial navigation system (INS) (acc. < 10 cm)	Outdoor
OpenLoris-Scene	2019	+	OptiTrack motion capture system / LIDAR SLAM (acc. < 10 cm)	Indoor
DRE dataset	2019	+	Downview camera and markers (acc. unknown)	Indoor
KAIST	2019	+ (rosbag player)	LIDAR SLAM (acc. unknown)	Outdoor
Malaga Urban	2014	-	GPS (low acc.)	Outdoor
Oxford RobotCar	2017	-	SPAN GNSS Inertial Navigation System (acc. < 10 cm)	Outdoor
RPNG OpenVINS	2019	+	GPS in outdoor scenes (low acc.)	Indoor / Outdoor
Segway DRIVE	2019	+	LIDAR SLAM (acc. < 10 cm)	Indoor
TUM RGB-D	2012	+	Motion capture system (acc. $\approx$ 1 mm)	Indoor
UMich NCLT	2015	+ (parser is provided)	Fused GPS/IMU/LIDAR (acc. $\approx$ 10 cm)	Outdoor
Zurich Urban MAV	2017	+ (parser is provided)	Pix4D SLAM (acc. unknown)	Outdoor

Table 4: Datasets' additional features.

Dataset name	Variable weather conditions	Dynamic objects	Memory consumption	Path length
EuRoC MAV	-	-	1-2.5 Gb	30-130 m
TUM VI	-	+	3-60 Gb	<20 km
KITTI	+	+	>20 Gb	<40 km
OpenLoris-Scene	-	-	6-33 Gb	Not mentioned
DRE dataset	-	Static, Low/High Dynamic	4-8 Gb	Not mentioned
KAIST	-	+	1-30 Gb	1-30 km
Malaga Urban	-	+	1-33 Gb	<36.8 km
Oxford RobotCar	+	+	10-500 Gb	Not mentioned
RPNG OpenVINS	-	-	1-2.6 Gb	27-105 m and 2.3, 7.4 km
Segway DRIVE	-	+	1-20 Gb	50 km
TUM RGB-D	-	+	0.2-2.5 Gb	0.4 km
UMich NCLT	+	+	80-110 Gb	1-7.5 km
Zurich Urban MAV	-	+	28 Gb	2 km

dataset. Dataset provides time-synchronized camera images with 1024x1024 resolution and IMU measurements. However, this dataset provides accurate pose ground truth from a motion capture system only at the start and end of the sequences. There is no ground truth data for the outdoor part of the sequence.

OpenLoris-Scene dataset [65] is a dataset that aims to help evaluate the maturity of SLAM and scene understanding algorithms for real-world deployment, by providing visual, inertial and odometry data recorded

with real robots in real scenes, and ground-truth robot trajectories acquired by motion capture system or high-resolution LiDARs. The important peculiarity of this dataset is quite hard real-life conditions with recordings of the same scenes in different lighting levels and with dynamic content.

TUM RGB-D SLAM Dataset [68] is a dataset that includes RGB-D data and ground-truth data with the goal to establish a novel benchmark for the evaluation of visual odometry and visual SLAM systems. The dataset contains the color and depth images of a Microsoft Kinect sensor and accelerometer data. The ground-truth trajectory was obtained from a high-accuracy motion-capture system.

KITTI [33] vision dataset was collected using an autonomous driving platform and scenes are captured by driving around the mid-size city of Karlsruhe, in rural areas and on highways. Dataset has a benchmark suite that includes odometry, object detection and tracking, road [29], stereo and flow benchmarks. The odometry

dataset was chosen for visual SLAM evaluation. This benchmark consists of 22 stereo sequences, saved in lossless *png* format and has a leaderboard. For this benchmark, it is possible to provide results using monocular or stereo visual odometry, laser-based SLAM or algorithms that combine visual and LIDAR information. The data was recorded using an eight-core i7 computer.

KAIST Urban Data Set [40] is another dataset collected using a car platform. The vehicle was equipped with two 2D and two 3D LiDARs to collect data on the surrounding environment. Additionally, the sensor suite included a stereo camera installed facing the front of the vehicle, GPS, VRS GPS, IMU, Fiber Optic Gyro (FOG) and altimeter. All sensor information was provided in a raw file format with timestamps. Three PCs were used to collect the data. The system clocks of the three PCs were periodically synchronized using the Chrony library [16]. Each PC used an i7 processor, 512 GB solid-state drive (SSD), and 32 GB DDR4 memory.

Oxford RobotCar Dataset [47] contains over 100 repetitions of a consistent route through Oxford, UK, captured over a period of over a year. The dataset captures many different combinations of weather, traffic and pedestrians, along with longer term changes such as construction and roadworks.

RPNG OpenVINS Dataset was developed by OpenVINS [34] authors. It includes ArUco datasets built using Synchronized Visual-Inertial Sensor System [53] in an indoor environment with ArUco markers. Also, this dataset includes two long sequences (2.3 and 7.4 km) build using ironides [77] visual-inertial sensor.

DRE dataset [75] is a dataset collected using the Redbot robot. It was made by DRE SLAM [75] authors for their algorithm evaluation. One of the key features of this dataset is the presence of odometry readings from wheel encoders. Moreover, the dataset includes scenes with low and high dynamic objects.

Segway DRIVE benchmark [39] includes datasets collected by Segway delivery robots deployed in real office buildings and shopping malls. Each robot was equipped with a global-shutter fisheye camera, a consumer-grade IMU synced to the camera on chip, two low-cost wheel encoders, and a removable high-precision lidar.

The Zurich Urban Micro Aerial Vehicle Dataset [48] was collected using camera equipped Micro Aerial Vehicle (MAV) flying within urban streets at low altitudes (i.e., 5-15 meters above the ground). The 2 km dataset consists of time-synchronized aerial high-resolution images, GPS and IMU sensor data, ground-level street view images, and ground truth data.

The Malaga Stereo and Laser Urban Dataset [7] was gathered entirely in urban scenarios with a car equipped

with several sensors, including one stereo camera (Bumblebee2) and five laser scanners. One distinctive feature of this dataset is high-resolution stereo images.

The University of Michigan North Campus Long-Term Vision and LIDAR Dataset [12] consists of omnidirectional imagery, 3D lidar, planar lidar, GPS, and proprioceptive sensors for odometry collected using a Segway robot.

The following Table 2 shows sensors used in the described earlier datasets. Datasets main and additional features are shown in the Tables 3 and 4. As it could be seen from the tables, there is no perfect dataset that matches different demands, such as indoor/outdoor data, changing conditions, dynamic content, several types of sensors, etc. As well as there is no unified way of data organisation. Several standards of ground truth formats exist (TUM, KITTY, EuRoC), but only this. It means that there is always a need to use several datasets and tune data formats for the algorithms which are itself demands special conditions for data.

### 3.4 Metrics

The result of the SLAM system is a trajectory and a map of the surrounding area. Despite the fact that it is theoretically possible to compare the resulting maps, in practice, obtaining reference maps is not a trivial task. In addition, different systems use different map formats, making this task even more difficult. For this reason, most often comparisons are made using the resulting trajectories.

However, it should be noted that a good trajectory does not necessarily mean a good map. For example, a map might be sparse with a low number of features or landmarks whereas a trajectory is precise. Small errors on the map, which have little effect on localization accuracy, can interfere with the normal functioning of the robot, for example, by blocking movement in a doorway or corridor.

Readings from additional sensors can be used to obtain reference trajectories. For outdoor datasets, high-precision GNSS systems are used in conjunction with the IMU sensor [32], while indoor optical tracking systems [69] can be used.

Relative Positional Error (RPE) [45] measures the trajectory drift over a fixed time interval  $\Delta$  in a reference frame, thus it does not accumulate previous errors. Additionally, many researchers tend to separate translational and rotational errors because it gives a more clear understanding of SLAM algorithm quality and it is not plausible to mix angles with poses (de-

grees with meters). For pose predictions of SLAM algorithm  $\mathbf{P}_1, \dots, \mathbf{P}_n \in \text{SE}(3)$ , and the ground-true poses  $\mathbf{Q}_1, \dots, \mathbf{Q}_n \in \text{SE}(3)$  the relative error for each point in time interval is defined as in [42]:

$$\mathbf{E}_i = (\mathbf{Q}_i^{-1} \mathbf{Q}_{i+\Delta})^{-1} (\mathbf{P}_i^{-1} \mathbf{P}_{i+\Delta})^{-1} \quad (2)$$

Thus, for a sequence of  $n$  measurements and their corresponding positions, there are  $m = n - \Delta$  relative errors. Using the error data, the root mean square error is calculated:

$$\text{RMS}(\mathbf{E}_{1:n}, \Delta) = \left( \frac{1}{m} \sum_{i=1}^m \|\mathbf{E}_i\|^2 \right)^{1/2} \quad (3)$$

In practice, it is useful to know Relative Rotational and Relative Translation errors for time interval  $\Delta$ . It allows estimating the quality of the SLAM algorithm for a particular situation without the influence of previous movements. Thus, rotational and translational parts can be calculated separately:

$$\text{RMS}(\mathbf{E}_{1:n}^{trans}, \Delta) = \left( \frac{1}{m} \sum_{i=1}^m \|\mathbf{E}_i^{trans}\|^2 \right)^{1/2} \quad (4)$$

$$\text{RMS}(\mathbf{E}_{1:n}^{rot}, \Delta) = \left( \frac{1}{m} \sum_{i=1}^m \|\mathbf{E}_i^{rot}\|^2 \right)^{1/2} \quad (5)$$

where  $\mathbf{E}_i^{trans}$  refers to the translation components of the Relative Pose Error and  $\mathbf{E}_i^{rot}$  refers to the rotational part.

In some cases, instead of the mean square, the median or means of other orders can be used. For example, the arithmetic mean will be less sensitive to anomalies than the root mean square, but more sensitive than the median.

For SLAM the global consistency of the resulting trajectory is also important, which can be estimated by calculating the difference between the corresponding estimated and reference poses (Absolute error) in a global frame. Since the frames of reference for the estimated and reference trajectories may differ from each other, the first step is to align them with each other. This can be done using the analytical solution provided in [36]. The solution finds a transformation  $\mathbf{S} \in \text{SE}(3)$  that minimizes the root-mean-square difference between  $\mathbf{P}_{1:n}$  and  $\mathbf{Q}_{1:n}$ . Using this transformation, the difference between the poses can be calculated as:

$$\mathbf{F}_i = \mathbf{Q}_i^{-1} \mathbf{S} \mathbf{P}_i$$

$$\text{RMS}(\mathbf{F}_{1:n}) = \left( \frac{1}{m} \sum_{i=1}^m \|\mathbf{F}_i\|^2 \right)^{1/2} \quad (6)$$

In practice, both metrics show a strong correlation with each other. But the main interest is in the comparison of different trajectories with one reference ground-truth trajectory for the whole length. Therefore, the more intuitive Absolute Error is used most often as well as in the results section. For the calculation of trajectories and metrics, the comparison EVO tool [35] was used. Exhaustive documentation, visualisation possibilities, and easy-to-use interface make it very convenient to compare a large number of results. Another possible way to track the performance is to use the KITTI Vision Benchmark Suite [33] web page. It allows seeing the translation and rotation error of an algorithm in a web table with more than a hundred different solutions, but it is limited to one dataset.

## 4 Results

Table 5: Results of SLAM algorithms tested on EUROC dataset

Framework	V1.01				MH.05			
	position(m)		rotation(deg)		position(m)		rotation(deg)	
	max	rms	max	rms	max	rms	max	rms
ORB-SLAM2	0.16	0.08	13.97	<i>4.70</i>	0.15	0.05	20.93	6.60
MapLab	0.30	0.15	10.03	6.13	1.13	0.59	2.99	1.27
LDSO	0.16	0.08	10.34	<i>4.68</i>	1.49	0.08	28.35	20.45
VINS-Mono	0.21	0.09	6.77	6.09	0.48	0.30	2.58	<i>0.79</i>
VINS-Fusion	0.24	0.07	10.19	5.60	0.40	0.19	4.81	1.87
OpenVSLAM	0.15	0.08	14.13	4.73	<i>0.14</i>	<i>0.04</i>	20.85	6.54
Basalt	<b>0.05</b>	<b>0.03</b>	<b>5.99</b>	5.36	0.17	0.09	<b>1.42</b>	<b>0.67</b>
Kimera	0.08	0.04	6.34	5.59	0.58	0.34	3.82	2.34
OpenVINS	0.11	0.05	<b>6.05</b>	5.45	0.48	0.16	3.36	1.30
ORB-SLAM3	<i>0.07</i>	<b>0.03</b>	6.39	5.94	<b>0.11</b>	<b>0.04</b>	<i>1.68</i>	0.93

Table 6: Results of SLAM algorithms tested on DRE dataset.

Framework	Low Dynamic (ST2)				High Dynamic (HD2)			
	position		rotation		position		rotation	
	max	rms	max	rms	max	rms	max	rms
ORB-SLAM2 mono	0.27	0.09	3.95	1.62	1.12	0.31	39.77	16.49
ORB-SLAM2 rgb-d	0.16	0.09	5.14	2.05	1.23	0.70	45.55	19.30
DRE-SLAM	<b>0.05</b>	<b>0.01</b>	<b>2.51</b>	<b>0.85</b>	<b>0.26</b>	<b>0.06</b>	<b>19.85</b>	<b>4.94</b>

The tables 5, 6, 7, 8, 9 represent the Absolute Pose Error metric (maximum and root mean squared values) for the considered algorithms. The metrics is both for position error in meters and for rotation error in degrees. The lowest error result among all algorithms is marked with bold green numbers and italic purple numbers as the second result. The first three tables represent results on popular and very common datasets in the robotics community. Not all algorithms are present in all tables. Absence means the inability to run on cer-

Table 7: Results of SLAM algorithms tested on TUM VI dataset.

Framework	Corridor 1				Magistrale 1				Room 1			
	position(m)		rotation(deg)		position(m)		rotation(deg)		position(m)		rotation(deg)	
	max	rms	max	rms	max	rms	max	rms	max	rms	max	rms
MapLab	0.78	0.27	4.18	2.85	2.72	0.88	13.92	6.37	0.27	0.14	6.05	2.10
VINS-Mono	1.74	0.44	1.40	0.47	0.35	0.24	3.64	1.06	0.12	0.05	3.79	1.74
OpenVSLAM	0.91	0.23	135.99	43.61	1.26	0.51	154.69	41.44	0.13	0.07	102.54	32.72
Basalt	1.34	0.33	4.25	2.32	1.91	0.68	5.28	1.47	0.23	0.09	5.92	1.25
OpenVINS	1.22	0.36	5.62	3.43	2.68	0.88	30.25	20.02	0.20	0.07	9.69	5.25
ORB-SLAM3	0.02	0.01	1.48	0.45	0.57	0.33	6.66	3.69	0.02	0.01	5.74	0.42

tain types of data, problems with parameters or algorithms themselves. Experiments on Open LORIS data and DRE dataset have been carried out for checking results in harsh conditions (sharp turns, textureless data, dynamic content). If a SLAM system could not be initialized or loses a solution it is marked as "X". The algorithms in the tables are placed in the order of their release date. The examples of obtained trajectories of algorithms on EuRoC MAV mh5 and TUM VI corridor 1 sequence are illustrated in Fig. 1.

Table 5 illustrates Basalt to be a winner on EuRoC V1\_01\_easy sequence, and ORB-SLAM3 and OpenVSLAM on MH\_05\_difficult. Worth reminding, "easy" EuRoC sequences differs from "difficult" one by jiggles and light conditions because of quadcopter data recording. ORB-SLAM3 and OpenVSLAM showed better results in worse conditions. In general, most of the algorithms show low errors and comparable results. Also, a much bigger error of LDSO can be noticed - the only Dense SLAM method which has been used.

Table 7 represents errors on TUM VI data sequence. Unfortunately, not all methods are capable to work with fish-eye data. The leader here is ORB-SLAM3 with errors, an order of magnitude less than others. It is worth noting, that this approach is 2 years newer than MapLab and VINS Mono - the second leaders, whereas all other algorithms in the table have similar release time. VINS Mono has two times higher error than MapLab on corridor\_1 sequence whereas it significantly wins on room1 and magistrale\_1 sequences. The length of magistrale\_1 sequence is much bigger (around 1 km) than others which makes it especially difficult for the algorithms. Even though OpenVSLAM shows good results on other sequences, here it is used to get lost, which led to huge rotational errors. The important peculiarity of this dataset is a handheld camera and lots of motion which makes IMU data especially important here.

Table 8 represents errors on KITTI dataset. Only the algorithms that could work with stereo data are provided. ORB-SLAM2 and OpenVSLAM show pretty

equal results with low errors (RMSE < 1 meter) on the majority of sequences. VINS Fusion and Basalt have several times worse results. Meanwhile, LDSO suffers much on KITTI sequences. As can be noticed, all methods have difficulties with sequence 02, which has an irregular shape and many smooth curves.

The summarized results of root mean squared errors for translation (in meters) and rotation (in degrees) obtained on three main datasets (TUM VI, EuRoC, KITTI) in Fig. 5 with different subset sequences. Histograms with the KITTI dataset are highlighted separately as they are based on traffic roads and one may be curious about the performance of different approaches in outdoor dynamic conditions. In terms of translation and rotation errors, ORB-SLAM 2/3 shows its superiority for all the sequences from each dataset. MapLab, OpenVSLAM, Basalt, and OpenVINS show mediocre results on TUM VI Magistrale 1 sequence, but OpenVSLAM has additional problems in rotation. VINS-Fusion and LDSO both obtain high errors on KITTI sequences, which impugns the usage of these approaches on public roads. Rotational errors on EUROC data are on a similar level of magnitude for all the algorithms except LDSO. This could be explained by the nature of data from a micro aerial vehicle. ORB-SLAM3, Basalt, Vins-Mono, and MapLab have the lowest rotational errors on the TUM VI sequences. Whereas OpenVSLAM is the only visual method on TUM VI that shows the highest error.

Experiments on the Open LORIS dataset (Table 9) aim to compare successful algorithms from previous runs that use data from different types of sensors. There are monocular rgb-d ORB-SLAM2, stereo VINS-Mono with IMU data, and OpenVSLAM with images from a stereo fish-eye camera but without IMU. The table shows that in general, all three algorithms have similar performance on several sequences. Looking at the details, it can be noticed that on the most crowded and dynamic sequences (cafe2 and home1), ORB-SLAM2 is losing trajectory at some moment. Moreover, monocular VINS mono overall has higher errors than stereo

Table 8: Results of SLAM algorithms tested on KITTI dataset.

Framework	00				02				05				06			
	position		rotation		position		rotation		position		rotation		position		rotation	
	max	rms	max	rms	max	rms	max	rms	max	rms	max	rms	max	rms	max	rms
ORB-SLAM2	2.82	<b>0.89</b>	<b>6.67</b>	<b>0.74</b>	<b>15.76</b>	6.60	4.37	1.47	<b>0.94</b>	<b>0.39</b>	<b>2.16</b>	<b>0.38</b>	<b>1.14</b>	<b>0.62</b>	<b>0.76</b>	<b>0.39</b>
LDSO	24.18	10.91	9.66	2.40	91.37	22.30	5.56	2.41	14.23	4.47	2.37	0.87	25.92	13.69	0.84	0.59
VINS-Fusion	13.80	5.20	7.93	3.22	47.90	18.17	9.31	4.78	16.78	4.79	7.49	3.02	7.14	2.50	6.93	4.81
OpenVSLAM	<b>2.41</b>	<b>0.88</b>	6.94	0.75	<b>12.88</b>	<b>5.53</b>	<b>4.18</b>	<b>1.20</b>	1.19	0.46	<b>2.15</b>	<b>0.38</b>	1.31	0.79	1.20	0.69
Basalt	5.09	3.38	7.25	0.88	19.49	9.11	4.88	1.72	4.96	2.48	2.30	0.86	4.00	2.11	3.17	2.78
ORB-SLAM3	<b>2.57</b>	0.90	<b>6.85</b>	<b>0.73</b>	16.27	<b>6.35</b>	<b>4.27</b>	<b>1.43</b>	<b>0.76</b>	<b>0.36</b>	2.18	<b>0.37</b>	<b>0.67</b>	<b>0.39</b>	<b>0.76</b>	<b>0.41</b>

Table 9: Results of SLAM algorithms tested on Open LORIS dataset.

Framework	office5				office7				cafe1				cafe2				home1			
	position		rotation		position		rotation		position		rotation		position		rotation		position		rotation	
	max	rms	max	rms	max	rms	max	rms	max	rms	max	rms	max	rms	max	rms	max	rms	max	rms
ORB-SLAM2	0.30	<b>0.21</b>	7.45	2.97	<b>0.09</b>	<b>0.04</b>	2.88	<b>1.45</b>	0.36	0.21	5.04	2.72	X	X	X	X	X	X	X	
VINS-Mono	0.32	0.22	5.83	3.38	0.40	0.12	3.75	2.42	1.31	0.42	5.40	3.47	0.75	<b>0.15</b>	<b>4.27</b>	<b>2.79</b>	1.09	0.55	16.01	8.86
OpenVSLAM	<b>0.28</b>	0.23	<b>5.12</b>	1.26	0.10	0.05	<b>2.86</b>	1.55	<b>0.14</b>	<b>0.10</b>	<b>3.35</b>	<b>0.77</b>	<b>0.42</b>	0.22	6.37	4.35	<b>0.50</b>	<b>0.36</b>	<b>9.36</b>	<b>2.30</b>

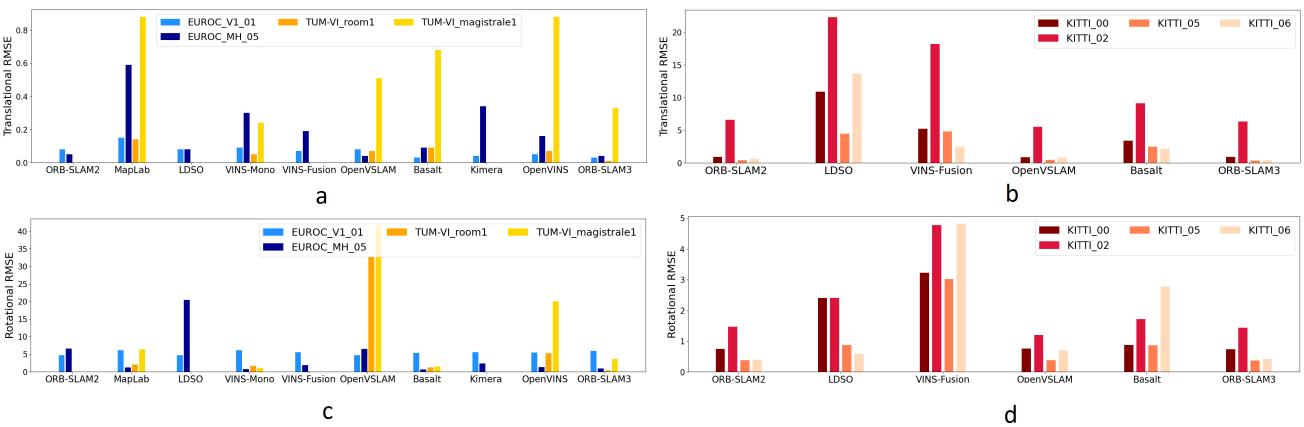


Fig. 5: Comparison of translational (a-b) &amp; rotational (c-d) root mean squared errors on EuRoC, TUM VI and KITTI datasets.

fish-eye OpenVSLAM. LDSO could not initialise these data. Also, the wheel encoder data is in an inappropriate data format for the DRE-SLAM.

In Table 6 the results of ORB-SLAM2 (mono and rgb-d versions) with DRE-SLAM on the DRE dataset with dynamic content are illustrated. It can be clearly seen that DRE SLAM [75] gives the best results both on low dynamic and high dynamic sequences in comparison with ORB-SLAM2.

Table 10 indicates CPU load and RAM memory usage. All algorithms were tested on the system with Intel Core i7 8565U CPU and 16 GB of RAM. It is worth mentioning that besides external modules of some approaches working in parallel the amount of required memory depends on many other aspects. A number of keyframes, amount of tracked visual features per keyframe, the type of a feature, and descriptor. The denser the map the more memory and computational time are needed to solve the problem. If the back-end of

the SLAM algorithm uses all the measurements in batch mode: Open VSLAM, ORB-SLAM 2/3, LDSO with g2o [45] solver or VINS-Mono, VINS-Fusion, Maplab, DRE with Ceres [2] solver, obviously, computational resources are limited, but this paradigm provides the best accuracy if not a significant amount of outliers present. However, the Bayes tree structure of the factor graph proposed in iSAM/iSAM2 [41] used in Kimera allows to significantly save resources due to incremental updates of Jacobian matrices with upcoming measurements. Only remote loop closures (between initial and last states) can nullify the advantage of a tree-based structure as all the poses have to be recalculated. The cheapest way to update trajectory and map is a fixed-lag smoothing approach with marginalization [38] used in Basalt which is a bit similar to filtering used in Open VINS. Unfortunately, it does not allow us to use all the measurements and might cause problems with remote loop closures. Recent algorithms use to solve a full

Table 10: CPU Load and Memory Usage.

Framework	Dataset	Memory RAM (Mb)			CPU Load (%)		
		Average	Max	Std	Average	Max	Std
Maplab	TUM VI corridor	526.5	721.7	36.1	85.2	523.2	34.9
VINS-Mono		87.0	92.1	1.1	68.5	118.8	36.2
Basalt		107.8	109.2	1.3	418.3	560.1	56.9
OpenVINS		1686.2	3329.1	952.5	85.2	523	34.9
Maplab	TUM VI outdoors	1587.5	4587.4	682.6	160.6	792.8	93.2
VINS-Mono		2851.0	4524.2	1285.9	182.5	306.7	94.2
OpenVSLAM		2639.6	4363.4	1071.3	230.6	310.5	38.8
Basalt		136.5	161.5	8.0	537.7	676.9	57.9
OpenVINS		6490.8	12819.5	3694.5	100.5	147.8	14.3
ORB-SLAM2	Euroc MH_05	705.9	869.9	102.7	187.2	249.8	37.0
LDSO		946.3	1082.1	102.9	160.1	263.9	50.6
Kimera		250.5	273.6	46.4	80.5	120.8	17.6
DRE-SLAM	DRE dataset hd2	5010.7	6288.1	1777.0	200.0	303.5	77.3

problem with all measurements in a parallel thread for saving resources while other threads are busy with current estimation. As table 10 illustrates, VINS-Mono demands fewer resources than others on the TUM VI corridor (indoor) environment but rather mediocre results on the outdoor sequence. Basalt shows a small need for memory both on indoor and outdoor sequences. OpenVINS consumes the smallest amount of CPU resources on TUM VI outdoors. All other algorithms on the outdoors sequence show significant demand for memory. The reason for that is the length of the sequence, leading to a huge amount of visual features. Kimera shows itself to be quite sufficient in memory consumption and CPU load on Euroc MH\_05 sequence. Results of ORB-SLAM2 and LDSO are comparable and quite demanding both in memory and CPU load. DRE-SLAM expectedly demands lots of resources even for a relatively short sequence. The possible reason is the use of several modules for a bunch of tasks.

## 5 Conclusions and recommendations

After a deep comparison of different algorithms, it could be concluded that ORB-SLAM2 and ORB-SLAM3 on average are the most reliable approaches for trajectory estimation. They show good results in different environments, sensor setups and even with dynamic objects. The reasons for that are a quality implementation and both fast and reliable visual features and the back-end part of the algorithm.

From a practical point of view, there is no perfect open-source out-of-the-box VIO/Visual SLAM solution for different environments and conditions. It is always needed to tune algorithm parameters for the particular

task, preprocess data, and solve dependency and performance issues.

**Visual-inertial approaches:** VINS-Mono, Basalt and OpenVINS benefit from using IMU data and show competitive results for versatile datasets. The drawback of visual-inertial fusion is the need for additional data sources, however, the majority of datasets provide this information. Meanwhile, one can mention that visual-only ORB-SLAM2, which was published five years ago, still can fight for 1-2 place. The same applies to the similar but refined OpenVSLAM, which got one of the best results on many datasets except TUM VI. Kimera represents a nice modular concept with great opportunities but low usability. Thus, it was hard to launch it on other datasets except for EuRoC. The only dense approach (LDSO) that had been tested showed far worse results compared to sparse algorithms. The modern ORB-SLAM3 shows its superiority in different setups and seems to be a perspective option to continue experiments. It tends to be a leader for the majority of datasets because of the welding map merging method [11] and novel place recognition technique. Finally, the full Bundle Adjustment problem is being solved [72]. Unfortunately, many dynamic SLAM approaches failed to run because of various issues. DRE-SLAM showed perfect results on dynamic data with the help of wheel encoders. But the problem is in the lack of wheel encoder data in datasets and it is an additional constraint to the setup during real-life data recording. The overall conclusion is that for different data and environments various approaches could show qualitative results and there is no striking leader. The most important is a correct experimental setup and the concept of limitations.

Another important point is that many SLAM solutions which have promising results in the paper but completely unusable GitHub repositories with a lack of support and examples only on the benchmark datasets. This refers to the reproducibility crisis which was stated in the introduction part of the paper.

Support and relevance of the implementation are important, as well as detailed instructions by the authors. Quickly updating packages cause dependency and repeatability issues even for a three-year-old solution.

The same is true for datasets. The absence of a complete dataset that could allow comparing different types of algorithms makes it difficult to understand which approach is the best one. Moreover, there is a need to unify data organisation both in datasets and in SLAM algorithms. Different formats of the ground truth data add a mess to work. For example, the KITTI format without timestamps makes it unusable for comparison of trajectories in other formats because of timestamps absence. Nevertheless, there is a significant choice of datasets suitable for comparing distinctive types of algorithms or specific environmental conditions.

It could be concluded that a user-friendly open-source slam solution is rather rare. One of them was OpenVSLAM. But during the paper writing, it has been deleted from the GitHub<sup>2</sup> due to the concerns on similarities of the source code to the ORB-SLAM2. Indeed, it can be noted that the resulting trajectories and errors are very similar in these two algorithms. Even though the National Institute of Advanced Industrial Science and Technology did not find the copyright infringement incidences, they decided to terminate the release of OpenVSLAM to avoid any risk of possible copyright issues.

Finally, this would be very convenient if every research work or SLAM approach had a Docker container to run and check the results described in the proposed method as it might solve reproducibility issues. even though the performance of an algorithm depends on computational resources, the Docker image allows using the same versions of the software so the difference will be only at a hardware level. To prove the practical consistency of an algorithm, it would be helpful if the latter had a ROS wrapper to check the proposed method on a real robot as a majority of the robotics community uses ROS. Also, this would be meaningful if robotics datasets had as many sensors as possible in order to check more approaches for a better understanding of benefits and disadvantages. Some methods could provide significantly better results on a dataset

<sup>2</sup> <https://github.com/xdspacelab/openvslam/wiki/Termination-of-the-release>

if there are sensor measurements needed for a visual SLAM approach.

It is also important to highlight several possible directions for future research in the area. For real-life applications, it could be reasonable to test algorithms on hardware-constrained microcomputing platforms. Another important aspect is the reliability of approaches for very long traversals, such as rides of self-driving cars. Lastly, it could be very beneficial to test the robustness of algorithms to outliers through a comparison of solvers and front-end parts of the algorithms.

## 6 Declarations

### Funding:

The authors declare that no funds, grants, or other support were received during the preparation of this manuscript.

### Conflicts of interest:

The authors have no relevant financial or non-financial interests to disclose, as well as no competing interests to declare that are relevant to the content of this article.

### Authors' contributions:

All authors contributed to the study conception and design. Material preparation, data collection and analysis were performed by Dinar Sharafutdinov, Mark Griguletskii, Pavel Kopanev, Mikhail Kurenkov and Aleksey Burkov. The review, text editing and article structure was made by Dzmitry Tsetserukou, Aleksei Gonnochenko and Gonzalo Ferrer.

## References

1. Agarwal, S., Mierle, K., Others: Ceres solver. <http://ceres-solver.org>
2. Agarwal, S., Mierle, K., Team, T.C.S.: Ceres Solver (2022). URL <https://github.com/ceres-solver/ceres-solver>
3. Antoni, R., Marcus, A., Yun, C., Luca, C.: Kimera: an open-source library for real-time metric-semantic localization and mapping. In: IEEE Intl. Conf. on Robotics and Automation (ICRA) (2020). URL <https://github.com/MIT-SPARK/Kimera>
4. Bărsan, I.A., Liu, P., Pollefeys, M., Geiger, A.: Robust dense mapping for large-scale dynamic environments. In: International Conference on Robotics and Automation (ICRA) (2018)
5. Bescos, B., Facil, J.M., Civera, J., Neira, J.: Dynaslam: Tracking, mapping, and inpainting in dynamic scenes. IEEE Robotics and Automation Letters **3**(4), 4076–4083 (2018). DOI 10.1109/Lra.2018.2860039. URL <http://dx.doi.org/10.1109/LRA.2018.2860039>
6. Biber, P., Straßer, W.: The normal distributions transform: A new approach to laser scan matching. pp. 2743 – 2748 vol.3 (2003). DOI 10.1109/IROS.2003.1249285

7. Blanco-Claraco, J.L., Moreno-Dueñas, F.A., González-Jiménez, J.: The málaga urban dataset: High-rate stereo and lidar in a realistic urban scenario. *Int. J. Rob. Res.* **33**(2), 207–214 (2014). DOI 10.1177/0278364913507326. URL <https://doi.org/10.1177/0278364913507326>
8. Bloesch, M., Czarnowski, J., Clark, R., Leutenegger, S., Davison, A.J.: CodeSLAM - Learning a Compact, Optimisable Representation for Dense Visual SLAM. In: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 2560–2568. IEEE (2018). DOI 10.1109/CVPR.2018.00271. URL <http://arxiv.org/abs/1804.00874><https://ieeexplore.ieee.org/document/8578369/>
9. Burri, M., Nikolic, J., Gohl, P., Schneider, T., Rehder, J., Omari, S., Achtelik, M.W., Siegwart, R.: The euroc micro aerial vehicle datasets. *The International Journal of Robotics Research* (2016). DOI 10.1177/0278364915620033. URL <http://ijr.sagepub.com/content/early/2016/01/21/0278364915620033.abstract>
10. Bustos, A.P., Chin, T.J., Eriksson, A., Reid, I.: Visual SLAM: Why Bundle Adjust? In: 2019 International Conference on Robotics and Automation (ICRA), vol. 2019-May, pp. 2385–2391. IEEE (2019). DOI 10.1109/ICRA.2019.8793749. URL <https://ieeexplore.ieee.org/document/8793749/>
11. Campos, C., Elvira, R., Gómez, J.J., Montiel, J.M.M., Tardós, J.D.: ORB-SLAM3: An accurate open-source library for visual, visual-inertial and multi-map SLAM. arXiv preprint arXiv:2007.11898 (2020)
12. Carlevaris-Bianco, N., Ushani, A.K., Eustice, R.M.: University of michigan north campus long-term vision and lidar dataset. *Int. J. Rob. Res.* **35**(9), 1023–1035 (2016). DOI 10.1177/0278364915614638. URL <https://doi.org/10.1177/0278364915614638>
13. Carbone, L., Kira, Z., Beall, C., Indelman, V., Dellaert, F.: Eliminating conditionally independent sets in factor graphs: A unifying perspective based on smart factors. IEEE International Conference on Robotics and Automation (ICRA) (2014). URL <https://ieeexplore.ieee.org/document/6907483>
14. Cesar, C., Luca, C., Henry, C., Yasir, L., Davide, S., Jose, N., Ian, R., J., L.J.: Past, present, and future of simultaneous localization and mapping: Toward the robust-perception age. *IEEE Transactions on Robotics* **32**(6), 1309–1332 (2016). DOI 10.1109/TRO.2016.2624754. URL <http://arxiv.org/abs/1606.05830><http://dx.doi.org/10.1109/TRO.2016.2624754>
15. Chen, C., Wang, B., Lu, C.X., Trigoni, N., Markham, A.: A Survey on Deep Learning for Localization and Mapping: Towards the Age of Spatial Machine Intelligence. arXiv (2020). URL <http://arxiv.org/abs/2006.12567>
16. Curnow, R.: Chrony. URL <https://chrony.tuxfamily.org>
17. Cvišić, I., Ćesić, J., Marković, I., Petrović, I.: SOFT-SLAM: Computationally efficient stereo visual simultaneous localization and mapping for autonomous unmanned aerial vehicles. *Journal of Field Robotics* **35**(4), 578–595 (2018). DOI 10.1002/rob.21762. URL <http://doi.wiley.com/10.1002/rob.21762>
18. Davison, A.J., Reid, I.D., Molton, N.D., Stasse, O.: MonoSLAM: Real-Time Single Camera SLAM. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **29**(6), 1052–1067 (2007). DOI 10.1109/TPAMI.2007.1049. URL <http://ieeexplore.ieee.org/document/4160954/>
19. Dellaert, F., Kaess, M.: Georgia tech smoothing and mapping (gtsam). IEEE International Conference on Robotics and Automation (ICRA) (2012). URL <https://smartech.gatech.edu/handle/1853/45226>
20. Delmerico, J., Scaramuzza, D.: A benchmark comparison of monocular visual-inertial odometry algorithms for flying robots. IEEE International Conference on Robotics and Automation (ICRA) (2018)
21. Dou, M., Khamis, S., Degtyarev, Y., Davidson, P., Fanello, S.R., Kowdle, A., Escolano, S.O., Rhemann, C., Kim, D., Taylor, J., et al.: Fusion4d: Real-time performance capture of challenging scenes. *ACM Transactions on Graphics (TOG)* **35**(4), 1–13 (2016)
22. Engel, J., Koltun, V., Cremers, D.: Direct sparse odometry. *IEEE transactions on pattern analysis and machine intelligence* **40**(3), 611–625 (2017)
23. Engel, J., Koltun, V., Cremers, D.: Direct sparse odometry. *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2018)
24. Engel, J., Schöps, T., Cremers, D.: Lsd-slam: Large-scale direct monocular slam. In: European conference on computer vision, pp. 834–849. Springer (2014)
25. Engel, J., Sturm, J., Cremers, D.: Semi-dense visual odometry for a monocular camera. In: Proceedings of the IEEE international conference on computer vision, pp. 1449–1456 (2013)
26. Fischler, M.A., Bolles, R.C.: Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography. In: Readings in Computer Vision, pp. 726–740. Elsevier (1987). DOI 10.1016/b978-0-08-051581-6.50070-2
27. Forster, C., Carlone, L., Dellaert, F., Scaramuzza, D.: On-manifold preintegration theory for fast and accurate visual-inertial navigation. *IEEE Trans. Robotics*: 33(1):1–21 (2016). URL <https://arxiv.org/abs/1512.02363>
28. Forster, C., Pizzoli, M., Scaramuzza, D.: Svo: Fast semi-direct monocular visual odometry. In: 2014 IEEE international conference on robotics and automation (ICRA), pp. 15–22. IEEE (2014)
29. Fritsch, J., Kuehnl, T., Geiger, A.: A new performance measure and evaluation benchmark for road detection algorithms. In: International Conference on Intelligent Transportation Systems (ITSC) (2013)
30. Galvez-López, D., Tardos, J.D.: Bags of Binary Words for Fast Place Recognition in Image Sequences. *IEEE Transactions on Robotics* **28**(5), 1188–1197 (2012). DOI 10.1109/TRO.2012.2197158. URL <http://ieeexplore.ieee.org/document/6202705/>
31. Gao, X., Wang, R., Demmel, N., Cremers, D.: LDSO: Direct Sparse Odometry with Loop Closure. In: 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 2198–2204. IEEE (2018). DOI 10.1109/IROS.2018.8593376. URL <https://ieeexplore.ieee.org/document/8593376/>
32. Geiger, A., Lenz, P., Stiller, C., Urtasun, R.: Vision meets robotics: The kitti dataset. *The International Journal of Robotics Research* **32**(11), 1231–1237 (2013)
33. Geiger, A., Lenz, P., Urtasun, R.: Are we ready for autonomous driving? the kitti vision benchmark suite. In: Conference on Computer Vision and Pattern Recognition (CVPR) (2012)
34. Geneva, P., Eckenhoff, K., Lee, W., Yang, Y., Huang, G.: Openvins: A research platform for visual-inertial estimation. In: 2020 IEEE International Conference on Robotics and Automation (ICRA), pp. 4666–4672 (2020). DOI 10.1109/ICRA40945.2020.9196524
35. Grupp, M.: evo: Python package for the evaluation of odometry and slam. <https://github.com/MichaelGrupp/evo> (2017)

36. Horn, B.K.: Closed-form solution of absolute orientation using unit quaternions. *Journal of the Optical Society of America A* **4**(4), 629–642 (1987). DOI 10.1364/JOSAA.4.000629. URL <http://people.csail.mit.edu/bkph/papers/Absolute-OCR.pdf>
37. Houseago, C., Bloesch, M., Leutenegger, S.: Ko-fusion: dense visual slam with tightly-coupled kinematic and odometric tracking. In: 2019 International Conference on Robotics and Automation (ICRA), pp. 4054–4060. IEEE (2019)
38. Hsiung, J., Hsiao, M., Westman, E., Valencia, R., Kaess, M.: Information sparsification in visual-inertial odometry. pp. 1146–1153 (2018). DOI 10.1109/IROS.2018.8594007
39. Huai, J., Qin, Y., Pang, F., Chen, Z.: Segway drive benchmark: Place recognition and slam data collected by a fleet of delivery robots (2019)
40. Jeong, J., Cho, Y., Shin, Y.S., Roh, H., Kim, A.: Complex urban dataset with multi-level sensors from highly diverse urban environments. *The International Journal of Robotics Research* p. 0278364919843996 (2019)
41. Kaess, M., Johannsson, H., Roberts, R., Ila, V., Leonard, J., Dellaert, F.: isam2: Incremental smoothing and mapping using the bayes tree. *International Journal of Robotic Research - IJRR* **31**, 216–235 (2012). DOI 10.1177/0278364911430419
42. Kasar, A.: Benchmarking and comparing popular visual slam algorithms. arXiv (2018)
43. Klein, G., Murray, D.: Parallel Tracking and Mapping for Small AR Workspaces. In: 2007 6th IEEE and ACM International Symposium on Mixed and Augmented Reality, pp. 1–10. IEEE (2007). DOI 10.1109/ISMAR.2007.4538852. URL <http://ieeexplore.ieee.org/document/4538852/>
44. Kopanov, P.: Slam-dockers. <https://github.com/KopanovPavel/SLAM-Dockers> (2021)
45. Kummerle, R., Grisetti, G., Strasdat, H., Konolige, K., Burgard, W.: G2o: A general framework for graph optimization. In: 2011 IEEE International Conference on Robotics and Automation, pp. 3607–3613. IEEE (2011). DOI 10.1109/ICRA.2011.5979949. URL <http://ieeexplore.ieee.org/document/5979949/>
46. Liu, Y., Fu, Y., Chen, F., Goossens, B., Tao, W., Zhao, H.: Datasets and Evaluation for Simultaneous Localization and Mapping Related Problems: A Comprehensive Survey. arXiv (2021). URL <http://arxiv.org/abs/2102.04036>
47. Maddern, W., Pascoe, G., Linegar, C., Newman, P.: 1 Year, 1000km: The Oxford RobotCar Dataset. *The International Journal of Robotics Research (IJRR)* **36**(1), 3–15 (2017). DOI 10.1177/0278364916679498. URL <http://dx.doi.org/10.1177/0278364916679498>
48. Majdik, A.L., Till, C., Scaramuzza, D.: The zurich urban micro aerial vehicle dataset. *Int. J. Rob. Res.* **36**(3), 269–273 (2017). DOI 10.1177/0278364917702237. URL <https://doi.org/10.1177/0278364917702237>
49. Mur-Artal, R., Montiel, J.M.M., Tardós, J.D.: ORB-SLAM: a versatile and accurate monocular SLAM system. *IEEE Transactions on Robotics* **31**(5), 1147–1163 (2015). DOI 10.1109/TRO.2015.2463671
50. Mur-Artal, R., Tardós, J.D.: ORB-SLAM2: an open-source SLAM system for monocular, stereo and RGB-D cameras. *IEEE Transactions on Robotics* **33**(5), 1255–1262 (2017). DOI 10.1109/TRO.2017.2705103
51. Newcombe, R.A., Davison, A.J., Izadi, S., Kohli, P., Hilliges, O., Shotton, J., Molyneaux, D., Hodges, S., Kim, D., Fitzgibbon, A.: KinectFusion: Real-time dense surface mapping and tracking. In: 2011 10th IEEE International Symposium on Mixed and Augmented Reality, pp. 127–136. IEEE (2011). DOI 10.1109/ISMAR.2011.6092378. URL <http://ieeexplore.ieee.org/document/6162880/>
52. Newcombe, R.A., Lovegrove, S.J., Davison, A.J.: DTAM: Dense tracking and mapping in real-time. In: 2011 International Conference on Computer Vision, pp. 2320–2327. IEEE (2011). DOI 10.1109/ICCV.2011.6126513. URL <http://ieeexplore.ieee.org/document/6126513/>
53. Nikolic, J., Rehder, J., Burri, M., Gohl, P., Leutenegger, S., Furgale, P.T., Siegwart, R.: A synchronized visual-inertial sensor system with fpga pre-processing for accurate real-time slam. In: 2014 IEEE International Conference on Robotics and Automation (ICRA), pp. 431–437 (2014). DOI 10.1109/ICRA.2014.6906892
54. Oleynikova, H., Taylor, Z., Fehr, M., Siegwart, R., Nieto, J.: Voxblox: Incremental 3d euclidean signed distance fields for on-board mav planning. arXiv (2017). DOI 10.1109/IROS.2017.8202315
55. Qin, T., Cao, S., Pan, J., Shen, S.: A general optimization-based framework for global pose estimation with multiple sensors (2019)
56. Qin, T., Li, P., Shen, S.: Vins-mono: A robust and versatile monocular visual-inertial state estimator. *IEEE Transactions on Robotics* **34**(4), 1004–1020 (2018). DOI 10.1109/TRO.2018.2853729
57. Qin, T., Shen, S.: Robust initialization of monocular visual-inertial estimation on aerial robots. In: 2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), vol. 2017-Septe, pp. 4225–4232. IEEE (2017). DOI 10.1109/IROS.2017.8206284. URL <https://ieeexplore.ieee.org/document/8206284/>
58. Qin, T., Shen, S.: Robust initialization of monocular visual-inertial estimation on aerial robots. In: 2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), vol. 2017-Septe, pp. 4225–4232. IEEE (2017). DOI 10.1109/IROS.2017.8206284. URL <https://ieeexplore.ieee.org/document/8206284/>
59. Rosinol, A., Sattler, T., Pollefeyt, M., Carlone, L.: Incremental visual-inertial 3d mesh generation with structural regularities. *IEEE Int. Conf. on Robotics and Automation (ICRA)* (2019). URL <https://arxiv.org/pdf/1903.01067>
60. Rublee, E., Rabaud, V., Konolige, K., Bradski, G.: ORB: An efficient alternative to SIFT or SURF. In: 2011 International Conference on Computer Vision, pp. 2564–2571. IEEE (2011). DOI 10.1109/ICCV.2011.6126544. URL <http://ieeexplore.ieee.org/document/6126544/>
61. Rünz, M., Agapito, L.: Co-fusion: Real-time segmentation, tracking and fusion of multiple objects. In: 2017 IEEE International Conference on Robotics and Automation (ICRA), pp. 4471–4478. IEEE (2017)
62. Schneider, T., Dymczyk, M., Fehr, M., Egger, K., Lynen, S., Gilitschenski, I., Siegwart, R.: Maplab: An open framework for research in visual-inertial mapping and localization. *IEEE Robotics and Automation Letters* **PP**, 1–1 (2018). DOI 10.1109/LRA.2018.2800113
63. Schubert, D., Demmel, N., von Stumberg, L., Usenko, V., Cremers, D.: Rolling-shutter modelling for direct visual-inertial odometry. arXiv preprint arXiv:1911.01015 (2019)
64. Schubert, D., Goll, T., Demmel, N., Usenko, V., Stückler, J., Cremers, D.: The tum vi benchmark for evaluating visual-inertial odometry. arXiv (2018). DOI 10.1109/IROS.2018.8593419

65. Shi, X., Li, D., Zhao, P., Tian, Q., Tian, Y., Long, Q., Zhu, C., Song, J., Qiao, F., Song, L., Guo, Y., Wang, Z., Zhang, Y., Qin, B., Yang, W., Wang, F., Chan, R.H.M., She, Q.: Are we ready for service robots? the OpenLORIS-Scene datasets for lifelong SLAM. In: 2020 International Conference on Robotics and Automation (ICRA), pp. 3139–3145 (2020)
66. Singh, R., Nagla, K.: Comparative analysis of range sensors for the robust autonomous navigation – a review. *Sensor Review ahead-of-print*, 17–41 (2019). DOI 10.1108/SR-01-2019-0029
67. Strasdat, H., Montiel, J.M., Davison, A.J.: Real-time monocular SLAM: Why filter? In: Proceedings - IEEE International Conference on Robotics and Automation, pp. 2657–2664 (2010). DOI 10.1109/ROBOT.2010.5509636
68. Sturm, J., Engelhard, N., Endres, F., Burgard, W., Cremers, D.: A benchmark for the evaluation of rgb-d slam systems. In: Proc. of the International Conference on Intelligent Robot Systems (IROS) (2012)
69. Sturm, J., Engelhard, N., Endres, F., Burgard, W., Cremers, D.: A benchmark for the evaluation of rgb-d slam systems. In: Intelligent Robots and Systems (IROS), 2012 IEEE/RSJ International Conference on, pp. 573–580. IEEE (2012)
70. Sumikura, S., Shibuya, M., Sakurada, K.: OpenVSLAM: A Versatile Visual SLAM Framework. In: Proceedings of the 27th ACM International Conference on Multimedia, MM '19, pp. 2292–2295. ACM, New York, NY, USA (2019). DOI 10.1145/3343031.3350539. URL <http://doi.acm.org/10.1145/3343031.3350539>
71. Tateno, K., Tombari, F., Laina, I., Navab, N.: CNN-SLAM: Real-time dense monocular SLAM with learned depth prediction. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) **2017-Janua**, 6565–6574 (2017). DOI 10.1109/CVPR.2017.695. URL <http://ieeexplore.ieee.org/document/8100178>/<http://arxiv.org/abs/1704.03489>
72. Triggs, B., McLauchlan, P.F., Hartley, R.I., Fitzgibbon, A.W.: Bundle adjustment — a modern synthesis. In: B. Triggs, A. Zisserman, R. Szeliski (eds.) *Vision Algorithms: Theory and Practice*, pp. 298–372. Springer Berlin Heidelberg, Berlin, Heidelberg (2000)
73. Usenko, V., Demmel, N., Schubert, D., Stückler, J., Cremers, D.: Visual-inertial mapping with non-linear factor recovery. arXiv (2019). DOI 10.1109/LRA.2019.2961227
74. Whelan, T., Salas-Moreno, R.F., Glocker, B., Davison, A.J., Leutenegger, S.: ElasticFusion: Real-time dense SLAM and light source estimation. *The International Journal of Robotics Research* **35**(14), 1697–1716 (2016). DOI 10.1177/0278364916669237. URL <http://journals.sagepub.com/doi/10.1177/0278364916669237>
75. Yang, D., Bi, S., Wang, W., Yuan, C., Qi, X., Cai, Y.: Dre-slam: Dynamic rgb-d encoder slam for a differential-drive robot. *Remote Sensing* **11**(4), 380 (2019)
76. Yu, C., Liu, Z., Liu, X.J., Xie, F., Yang, Y., Wei, Q., Fei, Q.: Ds-slam: A semantic visual slam towards dynamic environments. In: 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 1168–1174. IEEE (2018)
77. Zhang, Z., Liu, S., Tsai, G., Hu, H., Chu, C.C., Zheng, F.: Pirvs: An advanced visual-inertial slam system with flexible sensor fusion and hardware co-design (2017)