


# Under-canopy dataset for advancing simultaneous localization and mapping in agricultural robotics

The International Journal of  
Robotics Research  
2024, Vol. 43(6) 739–749  
© The Author(s) 2023  
Article reuse guidelines:  
[sagepub.com/journals-permissions](https://sagepub.com/journals-permissions)  
DOI: 10.1177/02783649231215372  
[journals.sagepub.com/home/ijr](https://journals.sagepub.com/home/ijr)  


Jose Cuaran , Andres Eduardo Baquero Velasquez, Mateus Valverde Gasparino, Naveen Kumar Uppalapati, Arun Narenthiran Sivakumar, Justin Wasserman, Muhammad Huzaifa, Sarita Adve and Girish Chowdhary

## Abstract

*Simultaneous localization and mapping (SLAM) has been an active research problem over recent decades. Many leading solutions are available that can achieve remarkable performance in environments with familiar structure, such as indoors and cities. However, our work shows that these leading systems fail in an agricultural setting, particularly in under the canopy navigation in the largest-in-acreage crops of the world: corn (Zea mays) and soybean (Glycine max). The presence of plenty of visual clutter due to leaves, varying illumination, and stark visual similarity makes these environments lose the familiar structure on which SLAM algorithms rely on. To advance SLAM in such unstructured agricultural environments, we present a comprehensive agricultural dataset. Our open dataset consists of stereo images, IMUs, wheel encoders, and GPS measurements continuously recorded from a mobile robot in corn and soybean fields across different growth stages. In addition, we present best-case benchmark results for several leading visual-inertial odometry and SLAM systems. Our data and benchmark clearly show that there is significant research promise in SLAM for agricultural settings. The dataset is available online at: <https://github.com/jrcuaran/terrasentia-dataset>.*

## Keywords

Agricultural robotics, agricultural dataset, visual odometry, simultaneous localization and mapping

Received 24 August 2023; Revised 27 October 2023; Accepted 27 October 2023

Senior Editor: Tim Barfoot

Associate Editor: Marija Popovic

## 1. Introduction

As the global population grows, there is an urgent need to boost agricultural productivity and sustainability. To this end, precision agricultural practices like under-canopy mechanical weeding, cover-crop planting, precision re-seeding, and high-throughput phenotyping are necessary. These practices can significantly enhance agricultural sustainability and reduce farmer costs through the reduction of agrochemical inputs. Particularly, high-throughput phenotyping enables measurements of key plant traits to speed up the development of more efficient agro-products such as seeds and chemicals. Small agricultural robots have shown significant potential for these tasks (R Shamshiri et al., 2018; Khanna et al., 2022). Both unmanned aerial and ground vehicles are powerful tools that can efficiently manage crop fields by performing tasks like monitoring, weeding, harvesting, and phenotyping (Oliveira et al., 2021). However, they face big challenges in agricultural environments because of their dynamic nature, uneven terrain, and varying illumination conditions, which lead to perception and control problems.

The localization of robots in agriculture and forestry has primarily relied on GPS measurements, sometimes fused with wheel odometry and IMU measurements (Oliveira et al., 2021). However, dense vegetation and plant canopy lower the accuracy of GPS measurements due to multi-path and signal losses (Aguilar et al., 2020). Therefore, there is a need for alternative methods that allow robots to localize themselves accurately. Visual odometry (VO) and Visual Simultaneous Localization and Mapping (V-SLAM) are viable alternatives in urban scenarios since they allow for the accurate estimation of the robot's pose. However, in agriculture, previous works have shown a poor performance of existing VO and V-SLAM algorithms due to repetitive

University of Illinois at Urbana-Champaign, Champaign, IL, USA

### Corresponding author:

Jose Cuaran, Field Robotics Engineering and Science Hub (FRESH), University of Illinois at Urbana-Champaign, 1304 W Pennsylvania Ave, Urbana, IL 61801, USA.  
Email: [jrc9@illinois.edu](mailto:jrc9@illinois.edu)

patterns, low texture, unstructured scenes, and variable lighting conditions (Cremona et al., 2022).

Further research is necessary to fully realize the potential of VO and V-SLAM algorithms in agriculture, and access to comprehensive and accurate agricultural datasets is crucial in facilitating this advancement. Although there have been some recently released agriculture datasets for robotics (Chebrolu et al., 2017; Pire et al., 2019; Imperoli et al., 2018), they are still limited to a few types of crops and growth stages. This paper introduces a comprehensive dataset collected by a ground robot in corn and soybean fields at different growth stages. It includes stereo images, IMU, GPS, and wheel encoder measurements. We provide multiple sequences with different terrains, occlusions, weather, and illumination conditions. Figure 1 shows the robot and sample images of these data. We hope that the high-quality data and the variability of this dataset accelerate the development of localization and mapping systems in agriculture. This paper also evaluates several visual-inertial odometry (VIO) and visual-inertial SLAM (VI-SLAM) systems on these data, in terms of accuracy, robustness, and computational efficiency.

This paper is organized as follows. Section 2 presents a compilation of related works, including publicly available datasets for localization and mapping in agriculture and some benchmarks. Section 3 describes the data acquisition system, the collection campaign, and the collected data. Section 4 presents a benchmark of VIO and VI-SLAM systems on some of the sequences of this dataset. Finally, the last section is devoted to discussion and conclusions.

## 2. Related work

### Previous datasets for VIO and VI-SLAM in agriculture:

There is a limited number of agricultural datasets publicly available for VIO and VI-SLAM. We can highlight the Rosario dataset (Pire et al., 2019), which includes six data sequences in a soybean field collected from a manually driven ground robot. This dataset consists of stereo images, GPS-RTK, IMU, and wheel encoder measurements. Even though the dataset includes dry and green crops and different vegetation densities, the data was recorded on only two dates, which prevents us from evaluating the effect of growth stages, weather, and varying illumination conditions. Moreover, due to the characteristics of the crop and the robot, the camera is always over the canopy, which prevents us from testing algorithms for under-canopy applications. Other limitations include low frame rate (15 fps), and the fact that most of the sequences start with the robot moving, making it impossible to test some algorithms that require stationary initialization (Pire et al., 2019).

Chebrolu et al. (2017) released an agricultural dataset for weed classification and localization and mapping in sugar beet fields. They used the BoniRob platform, equipped with multi-spectral sensors, RGB-D cameras, and 3D LiDAR. The data collection was done consistently three times per week for 3 months, leading to wide variability in the data



Figure 1. Terrasentia robot and sample images from the dataset.

sequences. This enables us to study the effect of growth stages, weather, and soil conditions on the performance of visual SLAM systems. However, since the cameras were mounted looking downwards, the application of these algorithms is limited to short crops, where over-canopy robots are feasible. In addition, the frame rate for both cameras is 1 fps, which is inappropriate for real-time applications. In contrast, LiDAR has an update rate of 20 Hz, making this dataset appropriate for LiDAR-based SLAM.

Kragh et al. (2017) published a dataset collected by a tractor mowing on a grass field. It consists of 2 h of data comprising stereo images, thermal images, radar, GPS, and others. This dataset is intended for obstacle detection, object tracking, localization, and mapping. It includes dynamic (humans) and static obstacles (mannequin dolls, rocks, barrels, buildings, vehicles, and vegetation). While these sequences are appropriate for object detection and tracking in open fields, they do not fully represent the challenges of agricultural fields regarding terrain traits, illumination, and occlusions. In addition to that, the sequences were collected on a single day, reducing variability. Finally, the low frame rate (10 Hz) is not ideal for real-time applications.

Unlike previous works, our dataset is suitable for evaluating localization and mapping algorithms under the canopy, which is a challenging task in agriculture due to occlusions and illumination variations. In addition, our data allows the evaluation of these algorithms at different growth stages, soil, and weather conditions. Finally, it includes high-resolution stereo images taken at 30 fps, synchronized with IMU measurements at 200 Hz, extending its application to VIO and VI-SLAM systems.

### Previous benchmarks of VIO and VI-SLAM systems in agriculture:

Recent benchmarks of VIO and VI-SLAM on agricultural datasets demonstrate the challenge of using robots in agricultural settings. Cremona et al. (2022) present one of the most comprehensive studies on the Rosario dataset. They tested several VIO and VI-SLAM systems, including Basalt, FLVIS, Kimera-VIO, OKVIS, ORB-SLAM3, REBVO, ROVIO, R-VIO, SVO 2.0, S-MSCKF, and VINS-Fusion. They found that S-MSCKF and ORB-

SLAM3 showed the best performance in terms of the absolute trajectory error (ATE). Even though the minimum mean value for ATE was 0.5 m obtained with MSCKF on sequences 3 and 4, this value is still high compared to their performance in indoor environments, where this metric can achieve values as low as a few centimeters (Servières et al., 2021; Campos et al., 2021). An error of 0.5 m in soybean fields is comparable to the row width and greater than the distance between plants, which limits the applications of these systems for high-precision agricultural tasks like phenotyping. They also found that only five out of 11 systems were able to complete the sequences. One of the issues they found is that some systems like Kimera-VIO, S-MSCKF, and R-VIO are sensitive to motion during the initialization. They also conclude that highly repetitive scenes, strong vibration, and moving leaves limit the performance of existing VIO systems. Similar results have been reported in different works evaluating VO and V-SLAM systems on the Rosario dataset, including ORB-SLAM2 and S-PTAM (Pire et al., 2019), S-PTAM, ORB-SLAM2, and S-MSCKF (Comelli et al., 2019).

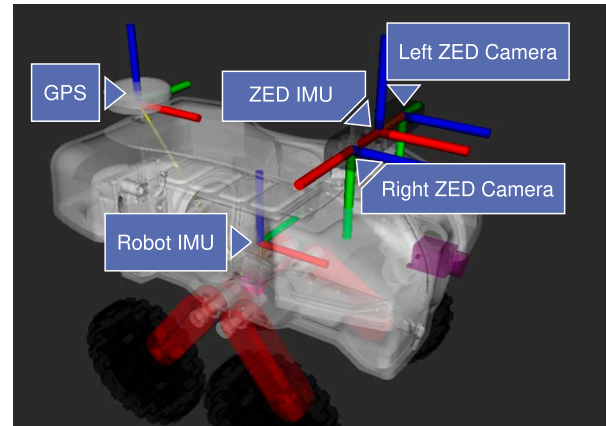
Similarly, Capua et al. (2018) tested stereo systems on fruit orchards. They ran LibViso, ORB-SLAM, and S-PTAM on three sequences collected from a ground robot, one in the summer and two in the winter. They found mean error values ranging from 1.9 m to 22 m. While ORB-SLAM showed the best performance, being able to close the loop in some sequences, LibViso did not complete the full trajectory, losing feature tracking early in the sequences. In addition, the mean errors were higher in the winter than in the summer, showing a significant influence of the seasons on the performance of VO and V-SLAM systems.

### 3. Materials and methods

#### 3.1. Acquisition system

**Robot platform:** The TerraSentia robot is a 4-wheeled skid-steering ground robot developed for data collection in outdoor environments. This robot is equipped with a stereo-inertial camera, a 6 DOF IMU, a GPS module, and wheel encoders. We used the Jetson AGX Orin computer to collect the sensor data and store it in a PCIe 4.0 SSD. This computer has an Nvidia GPU, which allows for efficient compression of images.

**Stereo-inertial camera:** We used the ZED2 stereo-inertial camera from StereoLabs. It includes a built-in IMU and returns rectified stereo images and a depth map synchronized with IMU measurements. It has a field of view of 120 deg. This camera is provided with a ROS wrapper that allows us to record data in two file formats, Rosbag (the standard ROS file format) and SVO (StereoLabs file format). For Rosbag files, we recorded images at a resolution of 832x468 px, which is suitable for VI-SLAM systems, while keeping files as small as possible. Thanks to efficient compression algorithms, SVO files are much more compact than Rosbag files, allowing us to record images at a resolution of 1280x720 px.



**Figure 2.** Coordinate frame system for each sensor mounted on the robot. X-axis (red), y-axis (green), and z-axis (blue).

#### 3.2. Sensor calibration

The camera's intrinsic, the rotation from the left camera's coordinate frame to other sensor's frames, and the noise parameters of the camera's IMU and the robot's IMU were estimated using the Kalibr calibration tool (Rehder et al., 2016). The translation of the GPS antenna with respect to the left camera frame was measured directly on the robot. Figure 2 shows the robot with the coordinate frames for each one of the sensors used during the data collection. Tables 1 and 2 show the summary of extrinsic and intrinsic parameters of the sensors. Further information related to sensor calibration is provided on the website of this dataset.

#### 3.3. Ground truth

All of the data sequences include GPS measurements, with 20% of them benefiting from RTK corrections. We mitigated the impact of the plant canopy on GPS accuracy by siting the RTK base station near the data collection field, thereby achieving a positional accuracy lower than 2 cm. The list of data sequences with reliable RTK corrections is available in the dataset's website.

#### 3.4. Data collection

Data was collected in the Illinois Autonomous Farm at the University of Illinois at Urbana-Champaign. Field corn is the main crop and was planted on different plots and dates. The collection campaign started in June and finished in September 2022. Several sequences were recorded twice per week, including fixed and random sequences. The former are those collected on the same corn rows during the whole season. These sequences, which include anywhere from one to four adjacent rows, are intended for evaluating the effect of different growth stages and weather conditions on the performance of navigation systems. In contrast, random sequences were collected each week on random rows and are intended to add more variability to the dataset in terms of weeds, terrain, and occlusions. Random

**Table 1.** Transformations from different coordinate frames to the left camera's coordinate frame.

Coordinate Frame	Rotation				Translation		
	$q_x$	$q_y$	$q_z$	$q_w$	$x(m)$	$y(m)$	$z(m)$
ZED IMU	0.49916	-0.49716	0.49954	0.50411	0.02184	-0.00014	-0.03020
Right camera	0.0	0.0	0.0	1.0	0.0	0.0	0.12
Robot IMU	-0.44643	0.53021	-0.52716	-0.49160	0.06000	0.24000	-0.01800
GPS	-	-	-	-	0.06000	-0.01000	-0.37000

**Table 2.** Intrinsic parameters of the stereo camera and noise parameters of the IMUs. The pinhole camera model with radial tangential distortion is considered for the ZED2 camera.

Distortion	Left camera	Right camera
K1	-0.03834527	-0.02184651
K2	0.02922888	0.01208824
r1	-0.00074899	-0.00099306
r2	0.00089755	0.00067692
Intrinsics	Left camera	Right camera
Fx	342.40289471	341.40687277
Fy	342.20510857	341.29153735
Cx	408.75398427	409.16655099
Cy	220.34050007	220.30732441
Noise	ZED IMU	Robot's IMU
Accel. noise density ( $m/s^2/\sqrt{Hz}$ )	0.00144514	0.40572120
Accel. random walk ( $m/s^3/\sqrt{Hz}$ )	0.00003447	0.00117599
Gyro. noise density ( $rad/s/\sqrt{Hz}$ )	0.00016381	0.00106726
Gyro. random walk ( $rad/s^2/\sqrt{Hz}$ )	0.00000698	0.00000768

sequences were also recorded in soybean, sweet corn, and sorghum fields, exhibiting different types of challenges in terms of texture and structure. In addition, we recorded data in the corridors between crop fields, as they are part of the environment where the robots need to navigate. Figure 3 shows the GPS tracks of some of these data sequences.

### 3.5. Data description

This dataset includes a total of 135 data sequences exhibiting occlusions, weeds, weather changes, and growth-stage variability as shown in Table 3. The dataset is organized in different folders, classified by crop type and collection date. We find Rosbag (\*.bag) and SVO (\*.svo) files within each folder. The Rosbag files comprise standard ROS messages for all the sensor measurements, including the stereo-inertial camera, IMU, GPS, and wheel encoders. It also includes the estimated ground-truth robot pose. Finally, it includes additional topics computed by the ZED software development kit (SDK) that we believed could be helpful for future research. Among these topics, we have a VIO topic and a camera pose topic (estimated by a VI-SLAM system). The SVO files comprise stereo images and sensor metadata. Further descriptions of this dataset and tutorials to get the data are included in the dataset's website.

## 4. Evaluation

We evaluated the performance of several VIO and VI-SLAM systems including ORB-SLAM3 (Campos et al., 2021), Openvins (Geneva et al., 2020), VINS-Fusion (Qin et al., 2019), SVO Pro (Forster et al., 2016), RTAB-Map (Labbé and Michaud, 2019), MSCKF (Sun et al., 2018), and ZED odometry. The latter is the odometry system operating on the ZED2 camera. The main characteristics of these systems are presented in Table 4. First, we should distinguish VIO from VI-SLAM systems. While VIO algorithms are intended for estimating the robot's ego-motion or local trajectory, VI-SLAM systems aim at continuously improving the estimation of the entire trajectory and the environment map. To this end, VI-SLAM systems rely on loop closure to detect previously visited places and add additional constraints to the global optimization step. This allows the system to fix the drift that pure VIO systems have (Fraundorfer and Scaramuzza, 2011). We also distinguish direct methods from indirect methods. The former are those in which different frames are matched by using pixel intensities, they commonly rely on reducing photometric error. In contrast, indirect methods rely on detecting features to be matched or tracked across different frames by reducing the reprojection error (Servières et al., 2021). Finally, these



VIO and VI-SLAM systems can be classified into filter-based and optimization-based methods. Filter-based methods estimate a state vector composed of the robot’s poses and landmarks’ positions, following prediction and update steps. The state vector is propagated at each time step, keeping only the last pose or a window with a limited number of the last poses. On the other hand, optimization-based methods perform global optimization by using bundle

adjustment or factor graph optimization. This is done over different poses along the whole trajectory, determined by consistently chosen keyframes (Strasdat et al., 2010).

These systems were evaluated in terms of accuracy, robustness, and computational efficiency. We ran the stereo-inertial implementation of each system using the zed IMU and camera, except for RTAB-Map which only relies on visual measurements. We configured all the systems with the same calibration parameters of the camera. However, as each method is sensitive to the IMU parameters, it was necessary to adjust them carefully for each of them by multiplying the original parameters by 10 or a hundred. Additional parameters specific to each method were also tuned to achieve high global consistency in a semi-structured outdoor environment, which is described below.

**Scenarios:** We tested these VIO and VI-SLAM methods on four data sequences exhibiting different levels of difficulty. Figure 4 displays some samples from these sequences. Scenario 1 consists of a double loop in a semi-structured environment close to a cornfield, including nearby static objects such as buildings and cars, which makes it typically easy for VIO and VI-SLAM systems to track features. Additionally, there are no occlusions or moving objects that could interfere with feature tracking. Notably, both low-texture elements (e.g., sky) and high-frequency texture components (e.g., grass) coexist within this scenario. Scenario 2 features a double loop in a corridor between corn and soybean fields. The main challenges in this environment are repetitive patterns and the absence of human-created objects. Scenario 3 comprises four consecutive corn rows at an early growth stage, where the primary challenges include harsh terrain leading to vibration and noisy measurements. Moving leaves are also highly sensitive to wind at this stage and can impact feature tracking. Furthermore, repetitive patterns in this scenario can lead to incorrect data associations and loop closures. Lastly, scenario 4 involves four adjacent corn rows at a late growth stage. This scenario shares the same challenges as scenario 3, along with variations in illumination and repetitive occlusions that can often blind the camera.

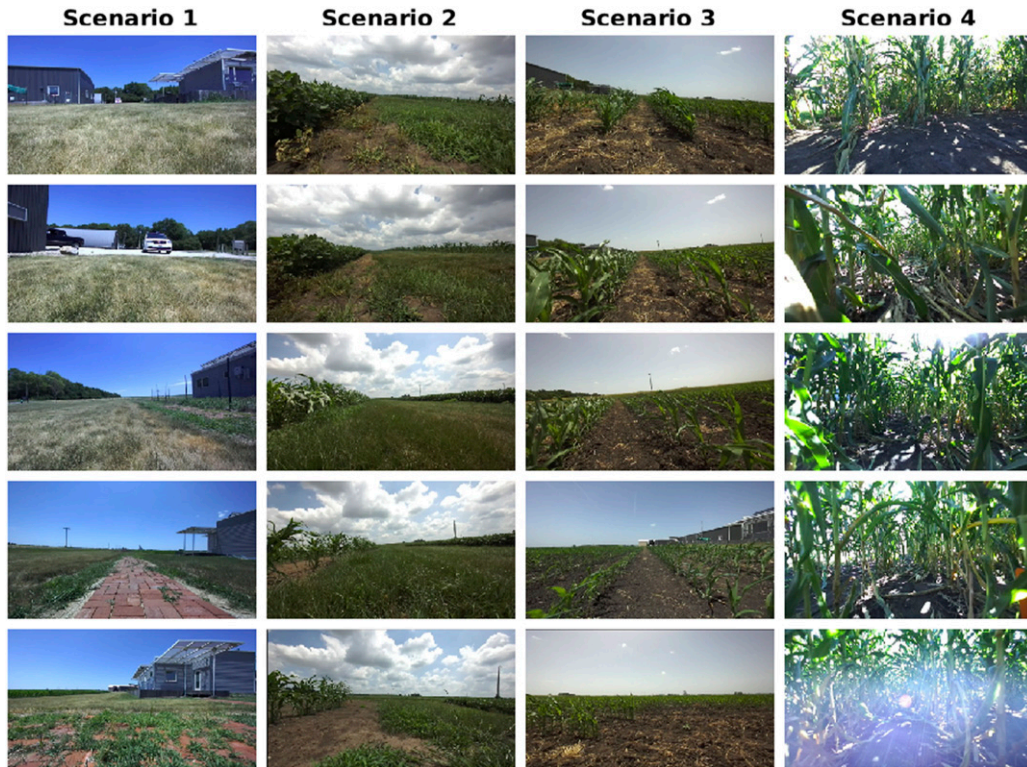


**Figure 3.** Satellite view of the farm with GPS tracks. The northern part (top) and the southern part (bottom) of the farm with a zoomed-in view of the main data sequences.

**Table 3.** Main characteristics of the data sequences.[illegible]

**Table 4.** VIO and VI-SLAM systems evaluated on the dataset.

System	Type	Map	Direct /indirect	Estimation approach	Loop closure
ORB-SLAM3 ( <a href="#">Campos et al., 2021</a> )	SLAM	Sparse	Indirect	Optimization	✓
VINS_Fusion ( <a href="#">Qin et al., 2019</a> )	VIO + SLAM	Sparse	Indirect	Optimization	✓
SVO Pro ( <a href="#">Forster et al., 2016</a> )	VIO + SLAM	Sparse	Semi-direct	Optimization	✓
RTAB-Map ( <a href="#">Labbe and Michaud, 2019</a> )	VO + SLAM	Dense	Indirect	Optimization	✓
Openvins ( <a href="#">Geneva et al., 2020</a> )	VIO	-	Indirect	Filtering	x
MSCKF ( <a href="#">Sun et al. 2018</a> )	VIO	-	Indirect	Filtering	x
ZED odometry	VIO	-	-	-	x

**Figure 4.** Sample images from different data sequences on which VIO and VI-SLAM systems were tested. First column: double loop in a semi-structured environment. Second column: double loop in a corridor between soybean and corn fields. Third column: four adjacent rows at an early growth stage. Fourth column: four adjacent rows at a late growth stage.

#### 4.1. Metrics

We utilized two metrics to evaluate the accuracy of VIO and VI-SLAM systems. The first is the absolute trajectory error (ATE), computed to evaluate the global consistency of the estimated trajectories. It computes the difference between points of the ground-truth trajectory and the estimated trajectory after aligning them. We also compute the relative pose error (RPE), which provides information about the local accuracy, particularly the translational and rotational drift per meter, which is more appropriate to evaluate the performance of VIO systems.

Since GPS measurements do not provide information on the robot's orientation, which is necessary to compute RPE,

we compute an approximation to the ground-truth robot pose by using an extended Kalman filter (EKF) and a moving horizon estimator (MHE) algorithm. The MHE is responsible for fusing GPS and control commands in a defined finite horizon while correcting for heading angle bias and estimating traction coefficients. Furthermore, the MHE is synchronized with the GPS at 5 Hz, and therefore slow. To improve the frequency of predictions, we augment the predicted states using the EKF. This uses gyroscope and accelerometer measurements for prediction and the MHE predictions for correction. More details about the implemented estimators are shown in ([Gasparino et al., 2022](#)).

## 5. Results

### 5.1. Accuracy

Table 5 shows the mean ATE and RPE values of the trajectories estimated by the VIO and VI-SLAM systems in each scenario (the best results are depicted in bold).

**Scenario 1:** All the SLAM systems were able to detect loop closures in scenario 1, achieving a mean ATE value smaller than 1.2 m in all cases, as shown in Table 5. That is

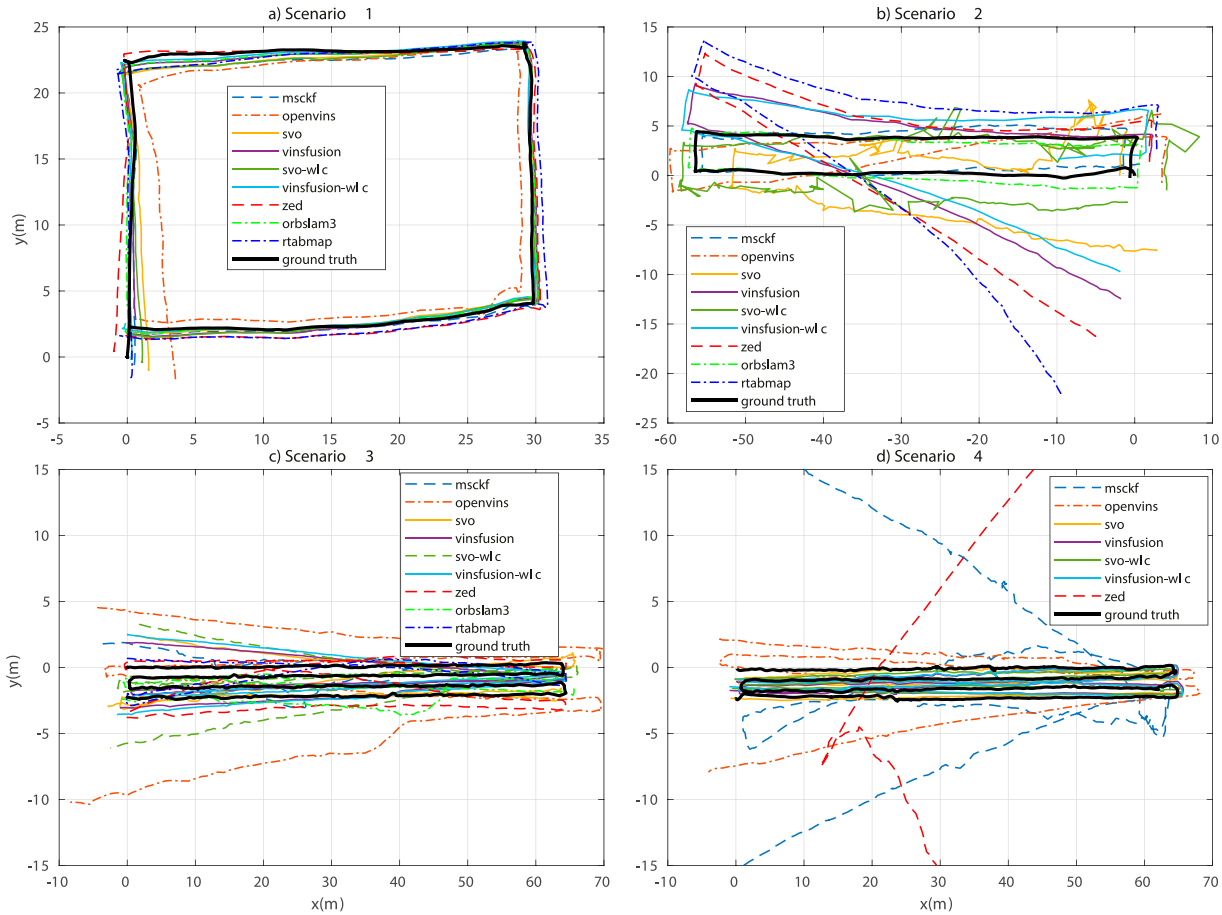
in line with the good consistency of the trajectories shown in Figure 5(a)). This was expected as this scenario is the easiest one, as it has neither occlusions nor close dynamic objects.

Despite the textureless of the sky, the frames have enough static and structured objects to compute distinctive features. However, similar ATE values were achieved by pure VIO systems. VINS-Fusion outperforms all the systems, achieving the smallest ATE value, therefore the best global consistency). Interestingly, it achieves better performance

**Table 5.** Accuracy values for VIO and VI-SLAM systems across scenarios. “×” for ATE and REP means that the system failed to complete the sequence, while “×” in “closed loop” denotes failure to perform loop closure. Best results are marked in bold.

System	ATE ↓ (SD)(m)	RPE ↓ (SD)(m)	RPE ↓ (SD)(deg)	Closed loop
<i>Scenario 1</i>				
RTAB-Map	0.681 (0.237)	0.155 (0.115)	3.952 (4.734)	✓
ORB-SLAM3	0.319 (0.142)	0.171 (0.114)	<b>3.614</b> (3.705)	✓
SVO-wlc	0.356 (0.204)	0.166 (0.105)	3.909 (4.544)	✓
VINS-Fusion-wlc	0.351 (0.113)	0.173 (0.141)	4.337 (5.427)	✓
MSCKF	0.518 (0.190)	0.179 (0.115)	4.233 (5.651)	-
Openvins	1.164 (0.348)	0.224 (0.167)	4.518 (6.002)	-
SVO	0.501 (0.222)	0.162 (0.106)	3.709 (3.995)	-
VINS-Fusion	<b>0.260</b> (0.130)	0.162 (0.116)	4.104 (4.921)	-
ZED odometry	0.592 (0.200)	<b>0.126</b> (0.112)	3.697 (4.136)	-
<i>Scenario 2</i>				
RTAB-Map	7.276 (5.229)	0.193 (0.106)	5.840 (4.661)	×
ORB-SLAM3	1.951 (1.036)	0.221 (0.186)	5.738 (4.860)	×
SVO-wlc	4.554 (3.321)	0.709 (0.605)	<b>4.436</b> (3.369)	×
VINS-Fusion-wlc	2.730 (1.851)	0.207 (0.103)	5.748 (4.451)	✓
MSCKF	<b>1.515</b> (0.419)	0.241 (0.123)	6.160 (5.379)	-
Openvins	2.749 (1.541)	0.224 (0.101)	5.765 (4.910)	-
SVO	5.063 (2.566)	0.708 (0.543)	4.881 (4.259)	-
VINS-Fusion	4.098 (2.823)	0.202 (0.115)	5.890 (4.892)	-
ZED odometry	5.798 (4.142)	<b>0.139</b> (0.092)	5.882 (4.770)	-
<i>Scenario 3</i>				
RTAB-Map	<b>1.127</b> (0.463)	<b>0.112</b> (0.056)	4.712 (3.456)	-
ORB-SLAM3	1.296 (0.712)	0.157 (0.104)	5.449 (3.014)	-
SVO-wlc	3.673 (1.578)	0.171 (0.109)	5.188 (2.888)	-
VINS-Fusion-wlc	1.613 (1.036)	0.128 (0.067)	5.354 (3.016)	-
MSCKF	2.190 (0.964)	0.392 (0.335)	<b>4.654</b> (2.737)	-
Openvins	4.392 (2.480)	0.216 (0.095)	4.923 (2.721)	-
SVO	4.860 (2.336)	0.176 (0.108)	5.331 (2.797)	-
VINS-Fusion	1.547 (0.888)	0.135 (0.069)	5.607 (3.193)	-
ZED odometry	2.821 (1.652)	0.139 (0.062)	5.287 (3.090)	-
<i>Scenario 4</i>				
RTAB-Map	×	×	×	-
ORB-SLAM3	×	×	×	-
SVO-wlc	0.707 (0.405)	0.170 (0.085)	4.838 (3.758)	-
VINS-Fusion-wlc	0.507 (0.240)	0.177 (0.093)	4.645 (3.516)	-
MSCKF	5.312 (4.620)	0.247 (0.194)	5.022 (3.661)	-
Openvins	2.426 (1.253)	0.190 (0.085)	<b>4.452</b> (3.373)	-
SVO	<b>0.477</b> (0.218)	<b>0.169</b> (0.085)	4.650 (3.785)	-
VINS-Fusion	0.597 (0.244)	0.175 (0.093)	4.499 (3.425)	-
ZED odometry	27.713 (10.800)	0.245 (0.672)	7.986 (6.501)	-





**Figure 5.** Estimated trajectories by VIO and VI-SLAM systems in different scenarios aligned to the ground truth with SE (3) Umeyama alignment. The VIO and VI-SLAM methods estimate consistent trajectories in scenario 1, leveraging the environment’s structure and rigidity. Conversely, in the remaining scenarios, significant drift is observed due to traits like repetitive patterns and moving leaves. For the sake of clarity, only the first loop is plotted for scenario 1 and scenario 2. Individual plots and ground-truth files can be accessed on the dataset’s website as supplementary materials.

than VINS-Fusion with loop closure (VINS-Fusion\_wlc), suggesting a small drift over time. Regarding RPE, all the methods exhibit close performance. The mean translational RPE ranges from 0.12 m to 0.22 m. These values are notably higher than those reported for VIO systems in indoor environments (e.g., Geneva et al., 2020) report translational RPE values as low as 0.05 m on the EuroMav dataset (Burri et al., 2016)). This difference could be attributed to the influence of distant points in outdoor scenes, which may not offer reliable translation information.

**Scenario 2:** Unlike scenario 1, only VINS-Fusion was able to detect loop closures in scenario 2. However, since it accomplished this later in the sequence, the estimated trajectory is not entirely consistent, as demonstrated in Figure 5(b). While all SLAM systems considered in this paper utilize the bag-of-words method (Gálvez-López and Tardós, 2012) for place recognition, known for its effectiveness and robustness in various environments, the repetitive patterns within the agricultural corridors can hinder accurate correspondence. Additionally, since this method necessitates a vocabulary of descriptors extracted from a set

of images, the vocabulary used by these methods may not align well with the characteristics of agricultural scenes. Table 5 illustrates mean ATE values reaching up to 7.3 m, reflecting trajectories with limited global consistency. Among the four SLAM systems, ORB-SLAM3 demonstrates the best performance, though its mean ATE value remains close to 2m, nearly six times larger than that achieved in scenario 1. As most SLAM systems fail to identify loop closures within this scenario, their performance aligns closely with that of pure VIO systems, both in terms of global and local consistency.

**Scenario 3:** In contrast to the preceding scenario, the rough soil and the presence of moving leaves along the corn rows in scenario 3 were anticipated to influence the performance of the VIO and VI-SLAM systems. Although both RTAB-Map and ORB-SLAM3 lost map tracking during the third row, all SLAM systems exhibited improved performance in this scenario (Figure 5(c)). Their achieved mean ATE values range from 1m to 3.7 m. Conversely, all VIO systems successfully completed the entire sequence without issue, showcasing translational drifts smaller than 0.4 m.



Overall, RTAB-Map, ORB-SLAM3, and VINS-Fusion demonstrate superior performance in terms of both global and local consistency.

**Scenario 4:** Scenario 4 becomes significantly more challenging than the previous one, as taller plants and longer leaves produce occlusions and affect illumination conditions. Certainly, RTAB-Map and ORB-SLAM3 could not track the robot pose in this environment. Since RTAB-Map relies on pure visual odometry only, occlusions become the main cause of failures. On the other hand, ORB-SLAM3 does consider inertial measurements for motion estimation, but it struggles to compute ORB features from the leaves. This is in agreement with previous works reporting the limitations of ORB-SLAM in low texture scenes (Merzlyakov and Macenski, 2021). Table 5 shows the accuracy metrics for the remaining systems. Overall, pure VIO methods exhibit similar performance as in the early growth stage in terms of translational and rotational RPE. This suggests that VIO are robust to occlusions and illumination variations. This can be explained by the fact that when robots get blinded,

they are still able to use inertial measurements to estimate their motion (this will be further discussed in the next section).

### 5.2. Robustness across growth stages

We evaluated the robustness of the VIO systems across different growth stages in a cornfield. To this end, we used data sequences of two random rows collected in consecutive weeks for 2 months. For each sequence, we measured the time that each VIO system was able to compute reasonable odometry values without interruption or divergence. It is assumed that the odometry diverges when the estimated velocity of the robot surpasses 2 m/s (physically impossible for the robot). We only evaluated VIO systems. The only exception is ORB-SLAM3, which does not have a switch to run pure visual odometry.

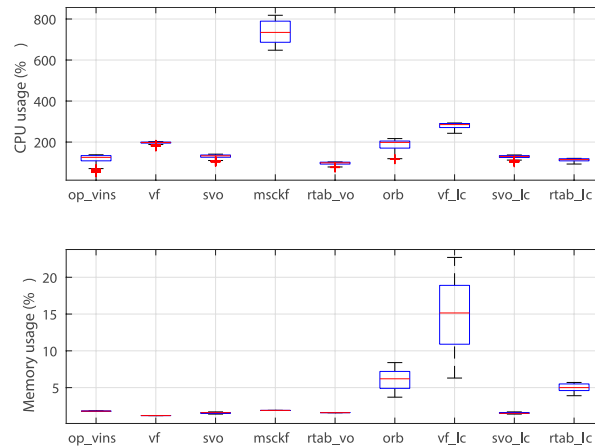
The main causes of failures include occlusions, wind, and illumination variations. Figure 6 shows some sample images of these failure cases. We can see in Table 6 that all the VIO systems can track more than 82% of the sequences across different growth stages. That confirms the robustness of these



**Figure 6.** Sample images of main causes of failure of VIO and SLAM systems. Wind is prevalent throughout all growth stages; however, crops in the early stages are more vulnerable, leading to failures of these systems, particularly during initialization. Occlusions predominantly occur during the middle growth stage, diminishing the quality of visual odometry, as occluded cameras force the systems to heavily depend on inertial measurements. Illumination variations are frequent and contingent upon weather conditions and time of day, resulting in feature tracking losses.

**Table 6.** Capability of tracking the robot trajectory in random corn rows at different growth stages (% of data sequence duration). Best results are marked in bold.

[illegible]



**Figure 7.** System efficiency measured by CPU and memory usage.

systems against occlusions, uneven terrain, and illumination changes. However, most of them failed on sequence June 27, in which wind caused the leaves to move faster than usual. Since these systems are not designed to work in dynamic environments, they lose tracking early in the sequence. On the other hand, RTAB-Map visual odometry only tracked an average of 8.4% of the total duration of the sequences, which is expected as pure visual odometry is sensitive to occlusions. Similarly, ORB-SLAM3 worked well during the early stages of the crops, but as the plants got taller, occlusions and low texture negatively impacted its performance. ORB-SLAM3 tries to create a new map after losing track. However, it loses tracking again as soon as new occlusions happen.

The poor robustness of ORB-SLAM3 in this dataset may be also explained by its feature-tracking approach. ORB-SLAM3 relies on descriptor matching, which may pose difficulties in scenarios with repetitive patterns and high-frequency textures objects (e.g., grass and soil) common in agricultural scenes. This is because each keypoint may have ambiguous correspondences from one frame to another, leading to incorrect associations. In contrast, methods such as VINS-Fusion, Openvins, and MSCKF are more robust in these types of scenarios, given their reliance on Lucas-Kanade tracking, a process iteratively resolved to minimize photometric errors. However, these failures could potentially be linked to refining parameter tuning, so further research is necessary.

### 5.3. Computational efficiency

Finally, we evaluated the computational efficiency of these systems in terms of CPU utilization and memory usage. We ran these systems on a laptop with an Intel Core-I7-11800H processor (which has 8 cores and operates at 2.30 GHz), and a 16 GB memory. Scenario 1 was used for the evaluation, as all the systems were able to complete this data sequence. We sampled the CPU utilization (in the percentage of a single core) and the resident memory (in the percentage of the total memory) at a sample rate of 1 Hz. Figure 7 shows these measurements. Overall, most of the VIO systems require less computational

resources than SLAM systems, since the last ones perform additional tasks like global optimization and loop closure, and allocate more memory for map features. In addition, RTAB-map odometry is the most efficient in terms of CPU load, which can be explained by being a pure visual-odometry algorithm. Finally, MSCKF shows the highest CPU load, which is even greater than those reported by previous works (Sun et al. 2018). Although the higher frame rate of our dataset (30 Hz) might be the main reason for that, finer tuning may be necessary to achieve the best performance of this system in new environments.

## 6. Conclusion

We have released a multisensor dataset suitable for the evaluation of VIO and VI-SLAM systems in agricultural fields. It comprises stereo images, inertial, encoders, and GPS measurements recorded weekly across the entire growth stages of corn crops. The wide variability of data in terms of weather, terrain, presence of weeds, illumination, occlusions, and wind disturbances makes this dataset one of the most comprehensive datasets publically available for localization and mapping in agriculture. A preliminary evaluation of VIO and SLAM algorithms showed that conventional SLAM systems do not work properly in agricultural scenarios, since feature tracking and loop closure fail due to factors like occlusions, repetitive patterns, and dynamic features. In contrast, VIO systems like VINS-Fusion, Openvins, and SVO Pro exhibit higher robustness and lower computational requirements, yet their drift over time is much higher than the reported in indoor environments.

## Acknowledgments

The authors would like to thank Pranav Jandamuri for his valuable support in the data collection process, and EarthSense Inc. for support with Terrasentia robot and data processing.

## Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

## Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported in part by NSF STTR Phase 2 #1951250 Agriculture and Food Research Initiative (AFRI) grant no. 2020-67021-32799/project accession no.1024178 from the USDA National Institute of Food and Agriculture: NSF/USDA National AI Institute: AIFARMS.

## ORCID iD

Jose Cuaran  <https://orcid.org/0000-0002-7281-9109>

## References

- Aguir AS, dos Santos FN, Cunha JB, et al. (2020) Localization and mapping for robots in agriculture and forestry: a survey. *Robotics* 9(4): 97. URL <https://www.mdpi.com/2218-6581/9/4/97>
- Burri M, Nikolic J, Gohl P, et al. (2016) The euroc micro aerial vehicle datasets. *The International Journal of Robotics Research* 35(10): 1157–1163.
- Campos C, Elvira R, Rodríguez JJG, et al. (2021) Orb-slam3: an accurate open-source library for visual, visual-inertial, and multimap slam. *IEEE Transactions on Robotics* 37(6): 1874–1890. URL <https://ieeexplore.ieee.org/document/9440682>
- Capua FR, Sansoni S and Moreyra ML (2018) Comparative analysis of visual-slam algorithms applied to fruit environments. In: *2018 Argentine Conference on Automatic Control (AADECA)*. Buenos Aires, Argentina: IEEE, pp. 1–6.
- Chebrolu N, Lottes P, Schaefer A, et al. (2017) Agricultural robot dataset for plant classification, localization and mapping on sugar beet fields. *The International Journal of Robotics Research* 36(10): 1045–1052. URL <https://journals.sagepub.com/doi/full/10.1177/0278364917720510>
- Comelli R, Pire T and Kofman E (2019) Evaluation of visual slam algorithms on agricultural dataset. *Reunión de trabajo en Procesamiento de la Información y Control* 1–6.
- Cremona J, Comelli R and Pire T (2022) Experimental evaluation of visual-inertial odometry systems for arable farming. *Journal of Field Robotics* 39: 1121–1135. URL <https://onlinelibrary.wiley.com/doi/10.1002/rob.22099>
- Forster C, Zhang Z, Gassner M, et al. (2016) Svo: semidirect visual odometry for monocular and multicamera systems. *IEEE Transactions on Robotics* 33(2): 249–265. URL <https://ieeexplore.ieee.org/document/7782863>
- Gálvez-López D and Tardós JD (2012) Bags of binary words for fast place recognition in image sequences. *IEEE Transactions on Robotics* 28(5): 1188–1197. DOI: [10.1109/TRO.2012.2197158](https://doi.org/10.1109/TRO.2012.2197158).
- Gasparino MV, Sivakumar AN, Liu Y, et al. (2022) Wayfast: navigation with predictive traversability in the field. *IEEE Robotics and Automation Letters* 7: 10651–10658. URL <https://ieeexplore.ieee.org/document/9839522/>
- Geneva P, Eckenhoff K, Lee W, et al. (2020) Openvins: a research platform for visual-inertial estimation. In: *2020 IEEE International Conference on Robotics and Automation (ICRA)*. Paris, France: IEEE, pp. 4666–4672. URL <https://ieeexplore.ieee.org/document/9196524>
- Imperoli M, Potena C, Nardi D, et al. (2018) An effective multi-cue positioning system for agricultural robotics. *IEEE Robotics and Automation Letters* 3(4): 3685–3692. URL <https://ieeexplore.ieee.org/document/8409986>
- Khanna M, Atallah SS, Kar S, et al. (2022) Digital transformation for a sustainable agriculture in the united states: opportunities and challenges. *Agricultural Economics* 53: 924–937. URL <https://onlinelibrary.wiley.com/doi/abs/10.1111/agec.12733>
- Kragh MF, Christiansen P, Laursen MS, et al. (2017) Fieldsafe: dataset for obstacle detection in agriculture. *Sensors* 17(11): 2579. URL <https://www.mdpi.com/1424-8220/17/11/2579>
- Labbe M and Michaud F (2019) Rtab-map as an open-source lidar and visual simultaneous localization and mapping library for large-scale and long-term online operation. *Journal of Field Robotics* 36(2): 416–446. URL <https://onlinelibrary.wiley.com/doi/abs/10.1002/rob.21831>
- Merzlyakov A and Macenski S (2021) A comparison of modern general-purpose visual slam approaches. In: *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. Prague, Czech Republic: IEEE, pp. 9190–9197. URL <https://ieeexplore.ieee.org/document/9636615>
- Oliveira LF, Moreira AP and Silva MF (2021) Advances in agriculture robotics: a state-of-the-art review and challenges ahead. *Robotics* 10(2): 52. URL <https://www.mdpi.com/2218-6581/10/2/52>
- Pire T, Mujica M, Civera J, et al. (2019) The rosario dataset: multisensor data for localization and mapping in agricultural environments. *The International Journal of Robotics Research* 38(6): 633–641. URL <https://journals.sagepub.com/doi/10.1177/0278364919841437>
- Qin T, Pan J, Cao S, et al. (2019) A general optimization-based framework for local odometry estimation with multiple sensors. *arXiv preprint arXiv:1901.03638* URL: <https://arxiv.org/abs/1901.03638>
- Scaramuzza D and Fraundorfer F (2011) Visual odometry: Part i: the first 30 years and fundamentals. *IEEE Robotics and Automation Magazine* 18(4): 80–92. URL <https://ieeexplore.ieee.org/document/6096039>
- Shamshiri R, Weltzien C, Hameed IA, et al. (2018) Research and development in agricultural robotics: a perspective of digital farming. *International Journal of Agricultural and Biological Engineering*. URL: <https://ijabe.org/index.php/ijabe/article/view/4278/0>
- Rehder J, Nikolic J, Schneider T, et al. (2016) Extending kalibr: calibrating the extrinsics of multiple imus and of individual axes. In: *2016 IEEE International Conference on Robotics and Automation (ICRA)*. Stockholm, Sweden: IEEE, pp. 4304–4311. URL <https://ieeexplore.ieee.org/document/7487628>
- Servières M, Renaudin V, Dupuis A, et al. (2021) Visual and visual-inertial slam: state of the art, classification, and experimental benchmarking. *Journal of Sensors* 2021, 1–26, URL <https://www.hindawi.com/journals/js/2021/2054828/>
- Strasdat H, Montiel J and Davison AJ (2010) Real-time monocular slam: why filter? In: *2010 IEEE International Conference on Robotics and Automation*. Anchorage, AK, USA: IEEE, pp. 2657–2664. URL <https://ieeexplore.ieee.org/document/5509636>
- Sun K, Mohta K, Pfrommer B, et al. (2018) Robust stereo visual inertial odometry for fast autonomous flight. *IEEE Robotics and Automation Letters* 3(2): 965–972. URL <https://ieeexplore.ieee.org/document/8258858>