

Virtual testbed for monocular visual navigation of small unmanned aircraft systems

Journal of Defense Modeling and Simulation: Applications, Methodology, Technology
2022, Vol. 19(3) 433–451
© The Author(s) 2020
DOI: 10.1177/1548512920954545
journals.sagepub.com/home/dms



Kyung Kim, Robert C Leishman¹, and Scott L Nykl¹

Abstract

Monocular visual navigation methods have seen significant advances in the last decade, recently producing several real-time solutions for autonomously navigating small unmanned aircraft systems without relying on the Global Positioning System (GPS). This is critical for military operations that may involve environments where GPS signals are degraded or denied. However, testing and comparing visual navigation algorithms remains a challenge since visual data is expensive to gather. Conducting flight tests in a virtual environment is an attractive solution prior to committing to outdoor testing. This work presents a virtual testbed for conducting simulated flight tests over real-world terrain and analyzing the real-time performance of visual navigation algorithms at 31 Hz. This tool was created to ultimately find a visual odometry algorithm appropriate for further GPS-denied navigation research on fixed-wing aircraft, even though all of the algorithms were designed for other modalities. This testbed was used to evaluate three current state-of-the-art, open-source monocular visual odometry algorithms on a fixed-wing platform: Direct Sparse Odometry, Semi-Direct Visual Odometry, and ORB-SLAM2 (with loop closures disabled).

Keywords

Monocular visual navigation, visual odometry, simulation, virtual testbed, small unmanned aircraft system

1 Introduction

The use of small unmanned aircraft systems (SUASs) has risen in recent years across commercial and military applications. A key interest area for further development is in expanding autonomous capabilities for these systems, of which navigation is a fundamental problem. Current navigation solutions suffer from heavy reliance on the Global Positioning System (GPS). This dependency presents a significant limitation, especially for critical operations that may be conducted in environments where GPS signals are degraded or actively denied. Therefore, alternative navigation solutions without GPS must be developed and monocular visual methods are a promising and attractive approach due to the relatively low weight and energy cost of a camera along with the wealth of information available in an image. Furthermore, visual cameras are a passive sensing technology making them generally more difficult to deny than active sensing technology, such as GPS.

A limitation in current monocular visual navigation research is that much of the research has focused on

ground-based vehicles or quadrotor platforms flying at low altitudes and speeds in either indoor or small, constrained outdoor environments. However, the military has a need for fixed-wing SUASs to conduct extended missions in large outdoor environments. The conditions experienced by a fixed-wing SUAS in this environment largely differ from those prior listed and the performance of visual navigation algorithms under these conditions must be explored.

Major obstacles inhibiting the ability to analyze and compare visual navigation algorithms include conducting flight tests and gathering accurate truth data. Flight tests

Air Force Institute of Technology, Wright-Patterson AFB, USA

Corresponding author:

Robert C Leishman, Air Force Institute of Technology, 2950 Hobson Way, WPAFB, OH 45433, USA.
Email: Robert.Leishman@afit.edu

can be a costly, time-consuming effort and some areas may be restricted or unavailable to fly over. Furthermore, gathering synchronized, accurate truth data at high frequencies is a difficult task in general and this is compounded for small flying vehicles, especially in outdoor environments.

This work addresses the issues listed above by developing a high-fidelity virtual environment for testing and analyzing visual navigation algorithms on both simulated imagery streams and real-world datasets. The virtual environment is then used to examine the performance of three current state-of-the-art visual odometry (VO) algorithms on a fixed-wing aircraft under high speeds, altitudes, and aggressive maneuvers. The algorithms tested are Direct Sparse Odometry (DSO),¹ Semi-Direct Visual Odometry (SVO),² and ORB-SLAM2 (with loop closures disabled).³ Results from simulation and real-world flight tests are also compared.

1.1 Camera calibration

A necessary pre-requisite for calculating motion from images is establishing a camera model. Camera models allow mapping between a feature's three-dimensional (3D) coordinates in the world and their corresponding two-dimensional (2D) image coordinates. The camera model common to all three algorithms is the pinhole model, which assumes that all light rays incident to an image pass through a single point.⁴ Cameras must be calibrated such that the image information conforms to the pinhole model assumption. The calibration finds the intrinsic parameters, including the focal lengths and image center coordinates and the lens distortion parameters. A popular method for calibrating cameras, and the method used in this work, involves taking pictures of a flat checkerboard with known dimensions from multiple positions and orientations.

1.2 Visual navigation categories

Visual navigation methods can largely be divided into two classes: VO and visual simultaneous localization and mapping (VSLAM). VO incrementally computes the relative motion of a vehicle or camera from a sequence of images and tracks a more localized environment.^{5,6} VSLAM falls under the larger umbrella of SLAM, which seeks to maintain a global map of the environment and a globally consistent trajectory, which are used to update and correct one other.^{7,8}

In addition, VO and VSLAM algorithms can both be categorized based on their image comparison approach, which use either feature-based or direct methods to compare two or more images and calculate the transformation describing the motion between the images. Feature-based approaches⁹ identify key points of interest throughout

images and extract feature vector representations that are invariant to changes in conditions, such as lighting, rotation, translation, and scale. Robust features are necessary to be able to calculate and find the same representations when viewed from different perspectives. Features are matched across images and the transformations between the images are calculated by optimizing the geometric error of the feature positions. Direct methods¹⁰ avoid the costly steps of extracting and matching features, and instead directly calculate the transformation that optimizes the photometric error, or the difference in pixel intensities, between images.

Finally, VO and VSLAM algorithms differ by the pixel density of images used to calculate motion. The pixel densities are sparse, semi-dense, and dense methods. Sparse methods use a few points throughout images, semi-dense methods use regions of pixels, and dense methods use every pixel in the images.

1.3 Related work

Parallel Tracking and Mapping (PTAM)¹¹ laid a foundation architecturally for modern VO and VSLAM algorithms. PTAM introduced the idea of splitting tracking and mapping into separate threads and using a keyframe-based approach to allow real-time bundle adjustments for higher overall performance and accuracy.

Dense Tracking and Mapping (DTAM)¹² is a fully dense and direct algorithm that utilizes a graphics processing unit (GPU) to achieve real-time performance. Using every pixel in an image, DTAM was shown to be significantly more robust under occlusions, camera blur, and defocus than other sparse, feature-based methods. However, most VO and VSLAM solutions for SUASs and other mobile robotics currently focus on implementing solely central processing unit (CPU)-based algorithms to minimize the required processing power.

Large-scale Direct SLAM (LSD-SLAM)¹³ is a semi-dense VSLAM algorithm using direct methods for pose estimation and a loop closure method for correcting trajectories when the camera appears to be viewing a scene that was previously observed.

DSO¹ is a fully direct and sparse VO algorithm. DSO performs optimizations over a sliding window of keyframes and optionally incorporates a photometric camera calibration in addition to a geometric calibration. A photometric calibration models the camera's exposure response and pixel attenuation (vignetting).¹⁴ This calibration can be beneficial for direct methods as they operate directly on pixel brightness intensities, but is unnecessary for feature-based methods since features are usually invariant to changes in lighting conditions.

ORB-SLAM is one of the leading feature-based VSLAM algorithms.¹⁵ ORB-SLAM uses three threads:

one each for tracking, mapping, and loop closing. One of the greatest strengths of ORB-SLAM is the efficiency of using the same features for tracking, re-localization, and loop detection. The authors improved upon the original work¹⁵ and created ORB-SLAM2.³ ORB-SLAM2 adds in a full bundle adjustment step over all keyframes and map points in a fourth thread. It also extends the algorithm to accommodate monocular, stereo, and/or RGB-D cameras.

SVO² is a hybrid VO algorithm that extracts features only on keyframes to initialize new map points, but uses a direct method to calculate motion for every frame. This allows SVO to achieve high frame rates. SVO has been used as a baseline algorithm in multiple works. REMODE¹⁶ is a monocular dense reconstruction algorithm that uses a GPU to achieve real-time performance. REMODE uses SVO to estimate the camera's pose. Follow on methods¹⁷ fused SVO outputs with inertial measurement unit (IMU) measurements to autonomously control a quadrotor and execute a trajectory through an onboard processor while providing a real-time dense 3D map of the traversed area using REMODE on an offboard computer. SVO 2.0¹⁸ extends the original SVO algorithm to large field-of-view (FOV) cameras, multi-camera systems, IMU incorporation, and the additional use of edges for feature alignment. However, the original SVO algorithm was used in this work since SVO 2.0 is currently only available as a pre-compiled binary and the source code is not publicly available.

A relative navigation architecture¹⁹ consists of a relative front-end running a multi-state constraint Kalman filter for state estimation and a global back-end maintaining a pose graph for optimizations and loop closures. The closest work to that described herein is a relative navigation method for fixed-wing unmanned aerial vehicles (UAVs).²⁰ The authors describe a simulation environment that evaluates Wheeler's relative navigation architecture. The simulated aircraft was flown over a cityscape image at 11 m/s and an altitude of 50 m. However, the algorithm was not yet fast enough to run under real-time constraints and the frame rate of the virtual aircraft camera was limited to three frames-per-second (fps).

Recently, new approaches to visual navigation have been explored involving deep-learning and artificial neural networks. DeepVO²¹ uses deep recurrent and convolutional neural networks to calculate pose estimates from image sequences. This method makes camera calibrations and geometric optimizations unnecessary and is able to recover absolute scale without external input, as all this is learned while training the neural net. However, these methods can require large amounts of data to train and make functional. Publicly available training data for flights of fixed-wing UAVs is limited. Several public VO and VSLAM datasets with synchronized truth data are available to test algorithm performances. The KITTI dataset²² contains data collected

from a station wagon driving through traffic at 10 fps. The TUM RGB-D dataset²³ consists of color and depth images from a Microsoft Kinect sensor at 30 fps in an indoor office environment from either a handheld configuration or on a wheeled robot. The TUM monoVO dataset¹⁴ demonstrates 50 sequences of monocular data exploring both indoor and outdoor environments from a handheld configuration. However, the ground truth was generated through LSD-SLAM and is not completely accurate. Others²⁴ presented the European Robotics Challenge Micro Aerial Vehicle (MAV) dataset with a hexrotor flying in an indoor industrial environment. Ground truth data were recorded with either a laser tracking system at 20 Hz with millimeter accuracy or a motion capture system at 100 Hz. Finally, the Zurich Urban MAV dataset²⁵ is based on a quadrotor flying outdoors over the streets of Zurich, Switzerland, at altitudes of 5–15 m over a 2 km trajectory. However, ground truth data were generated using appearance-based topological localization and VSLAM algorithms. Each dataset described above contains inherent limitations and, most importantly, are largely inapplicable to fixed-wing SUASs flying in large outdoor environments.

Simulation environments for VO are becoming more common. The increased interest is primarily due to the large data requirements for training deep-learning approaches for visual processing, for example Gaidon et al.²⁶ Researchers have found that even large, real-world datasets (e.g., KITTI²²) are insufficient to satisfy the insatiable needs of training a deep neural network. There are a few approaches in the literature that are similar to the proposed approach, but none are focused on the needs and particularities of fixed-wing UAVs.

The approach outlined by Sayre-McCord et al.²⁷ presents a methodology for creating impressive photo-realistic image generation that is then processed by a VO pipeline. Estimates are then used within the control loop of the UAV. The approach is geared toward rotorcraft UAVs in small environments.

Microsoft has released an open-source simulation environment, known as AirSim,²⁸ built using the Unity game engine. The environment supports driving vehicles and multirotor UAVs and can stream out imagery and point clouds from virtual cameras/scanners mounted on platforms. The platform currently does not support fixed-wing UAVs. The Unity engine has been used in other efforts for similar purposes.²⁹

Gazebo³⁰ is the go-to simulation environment for robotics research, supported by the ROS middleware. Several packages have been built to support multirotor unmanned vehicles, such as RotorS,³¹ but the environment is lacking in support for fixed-wing vehicles.

This work presents a novel virtual testbed for conducting simulated flight tests over real-world terrain and for playing back previously recorded flight data. The real-time

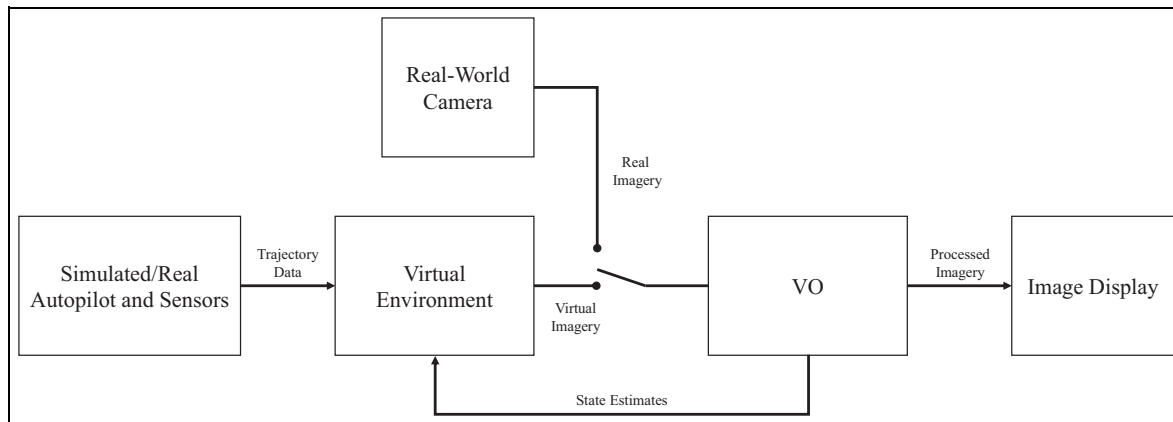


Figure 1. Navigation testbed system. The navigation testbed is a flexible and modular virtual environment that allows both simulated and real-world flight profiles to run in real-time over Global Positioning System-terrain-mapped satellite imagery. Estimated trajectories from visual odometry algorithms are projected in real-time and can be compared to one another and to truth. The simulator is also able to record and play back trajectories.

performance of visual navigation algorithms can then be analyzed and compared simultaneously to the performance of other algorithms within the same virtual testbed. The virtual environment was built using the AftBurner engine, a cross-platform visualization engine.³² The engine has been used successfully for testing stereo computer vision techniques in simulation for automated aerial refueling applications³³ and for testing structure from motion.^{34–36}

2 Simulation system

This work makes two contributions by using a novel collection of existing open-source tools combined with a couple of new capabilities developed as part of this effort. The first contribution is an ability to generate synthetic, physically correct trajectories and imagery for fixed-wing UAV vehicles. The second is a pluggable framework optimized for testing and evaluating VO algorithms, which features a 3D visualization capability to compare algorithm performance. This section will document the simulation system. First an overview of the framework will be presented and then each major component will be discussed in some detail. We outline below the sources for the different existing open-source components and the novel additions that combine to form the entire system.

2.1 System overview

The navigation testbed is a flexible and modular virtual environment that allows both simulated and real-world flight profiles to run in real-time over GPS-terrain-mapped satellite imagery. Estimated trajectories from VO algorithms are projected in real-time and can be compared to

one another and to truth. The simulator is also able to record and play back trajectories.

Figure 1 illustrates a high-level flow diagram of the system and Figure 2 shows the components of the overall navigation testbed system. The terrain is composed of satellite imagery augmented with elevation data. A UAV's trajectory data is generated through a simulated or real-world autopilot. Corresponding sensor output can be projected into the virtual environment. This projection enables one to visualize what the UAV is actually sensing and place the data in the context of the UAV's current geographical position and orientation.

The virtual environment produces imagery from a virtual camera attached to the UAV flying a desired trajectory. This imagery is then used as input for a VO algorithm. Alternatively, the VO algorithms can be used to process imagery streams from a real-world camera. The VO results are then fed back into the virtual environment to allow live visualization and analysis of the VO's six degree-of-freedom performance. Each of the VO algorithms also provides the processed state of the input imagery showing information such as which pixels or features were selected and used. This can be displayed through a separate node.

2.2 Autopilot interface

The ArduPilot (<https://ardupilot.org/dev/index.html>) open-source autopilot software is used to control the aircraft in both the virtual environment and the real world. ArduPilot uses an extended Kalman filter (EKF) to fuse sensor input and provide low-level control for various autonomous vehicles, including rovers, submarines, copters, and planes. Raw and fused sensor data can be read from the ArduPilot.

ArduPilot provides a Software in the Loop (SITL) program that allows ArduPilot software to be run on a

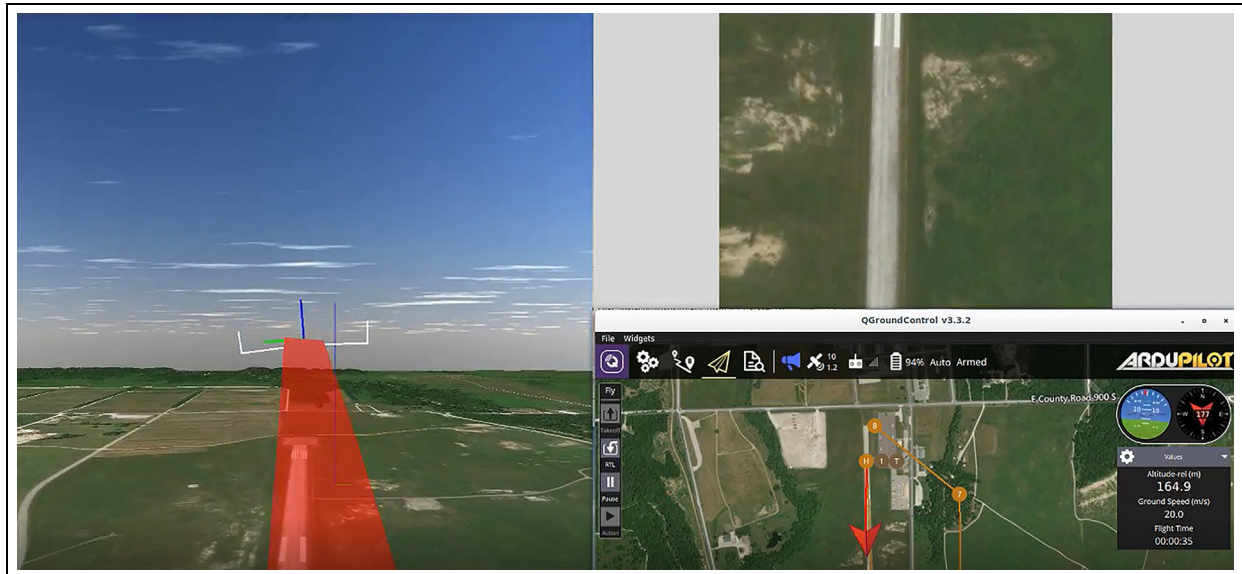


Figure 2. Simulator view when using completely simulated information. The left-hand window shows the virtual environment, with georegistered imagery draped on digital elevation model (DEM) tiles. The top right-hand window shows the live image feed from the virtual camera attached to the bottom of the aircraft. The bottom right-hand window shows the ground control system used to plan flights and control the aircraft.

computer and simulate virtual autopilots for vehicles. This is used to generate simulated trajectories for a fixed-wing aircraft exhibiting realistic flight dynamics.

A ground control station (GCS) program is required to give commands and communicate with the ArduPilot. The GCS used for controlling the simulated aircraft is QGroundControl since it is available on Linux, allowing simultaneous execution of all modules on the same laptop computer. The GCS used for controlling the real-world aircraft is Mission Planner, since that is the standard GCS used in the Air Force Institute of Technology (AFIT) Autonomy and Navigation Technology (ANT) Center.

The ArduPilot uses the MAVLink protocol to transmit and receive data. This data can be streamed or logged. The simulation system can connect directly to a real or SITL autopilot or load logged autopilot data.

2.3 Virtual environment

The virtual environment box in Figure 2 is actually composed of several components. The primary component is the AftBurner Engine, which then loads several other modules as plugins. These components are detailed below. The novel software contributions of this work were created as plugins to the AftBurner engine.

2.4 AftBurner engine

All virtual environment components are built using the AftBurner engine, a cross-platform visualization engine

written in C++ and the successor to the STEAMiE educational game engine.³² The visualization engine contains sub-modules for creating and reading virtual camera data as well as generating scaled real-world terrain models using United States Geological Survey (USGS) elevation data and satellite imagery. The AftBurner engine also contains formulations for modeling the World Geodetic System (WGS) 84 model to accurately visualize GPS coordinates. Figure 3 shows an example terrain model of the Grand Canyon created through the AftBurner engine mapped against GPS coordinates and USGS elevation data. Limiting factors include the available resolution of terrain and satellite imagery and the image distortion incurred from mapping a 2D overhead image onto a 3D model. This distortion is more pronounced in extreme elevation changes, as shown on the cliffs and walls of the virtual Grand Canyon. We did not test over areas of extreme elevation changes in this work. The smearing and distortion present in Figure 3 would cause accuracy issues with the VO algorithms tested.

The AftBurner engine is the core of the virtual environment and most of the capabilities used in this work are off the shelf from the available plugins. However, the following are new additions, created for this work.

2.5 Camera calibration

A camera calibration module was constructed to extract the intrinsic parameters of a monocular camera using the pinhole camera model within the AftBurner Engine. The module accepts imagery streams from a camera viewing a

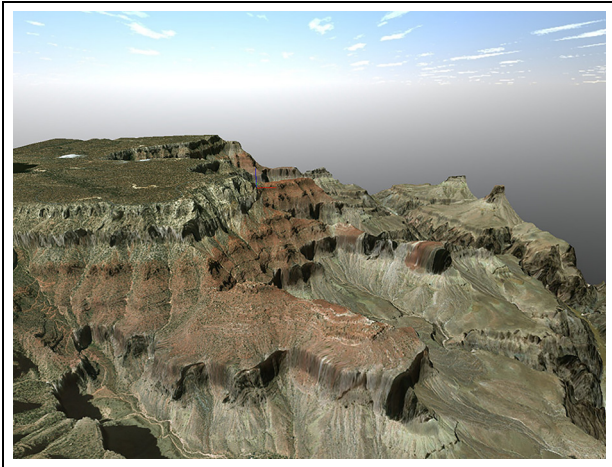


Figure 3. Virtual Grand Canyon terrain as an example of the capability of the AftBurner engine. Note that the distortion caused by the large changes in elevation would have consequences for the visual odometry algorithms and, consequently, such terrain was not employed for this study.

checkerboard as input and works interchangeably with virtual camera imagery as well as real-world camera imagery streams. Figure 4 shows an example of using the camera calibration module with a virtual camera. The left-hand window shows the perspective of the camera in the virtual environment. The right-hand windows show the camera calibration module, which presents both a live stream of the input imagery received and the last image captured for calibration. The calibration results are then used in the VO algorithms.

2.6 Trajectory viewer

A trajectory viewer module is used to view multiple trajectories simultaneously. This plugin was developed as part of this work and is original. This module is primarily for playback of a single truth trajectory and multiple estimated trajectories from VO algorithms applied to that truth trajectory. Trajectory playback is synchronized across all trajectories, can be skipped to any point in time, and sped up or slowed down. Three-dimensional plots of the trajectory and 2D plots for x , y , z , roll, pitch, and yaw data can also be generated. This allows detailed analysis and visualization of algorithm performances in all six degrees of freedom and to scale. Figure 5 shows an example of the trajectory viewer displaying multiple trajectories.

A current limitation of the trajectory viewer is that it can only load one dataset during each execution. Therefore, an additional trajectory plot viewer module is provided to allow analysis and plotting of multiple datasets in one process. It also provides more extensive plotting capabilities to be able to inspect individual plots in more detail as well as calculate and print out the trajectory position and rotation errors. Figure 6 shows the controls for the trajectory plot viewer. The left-most window pane allows the selection of directories containing truth and estimated trajectories for a given flight. Selecting a directory auto-populates the next window pane with the trajectories to allow selective plotting. The truth trajectory file must also be identified since it is used to synchronize the estimated trajectory files to ensure that pose plots are generated against the same flight times. The next window

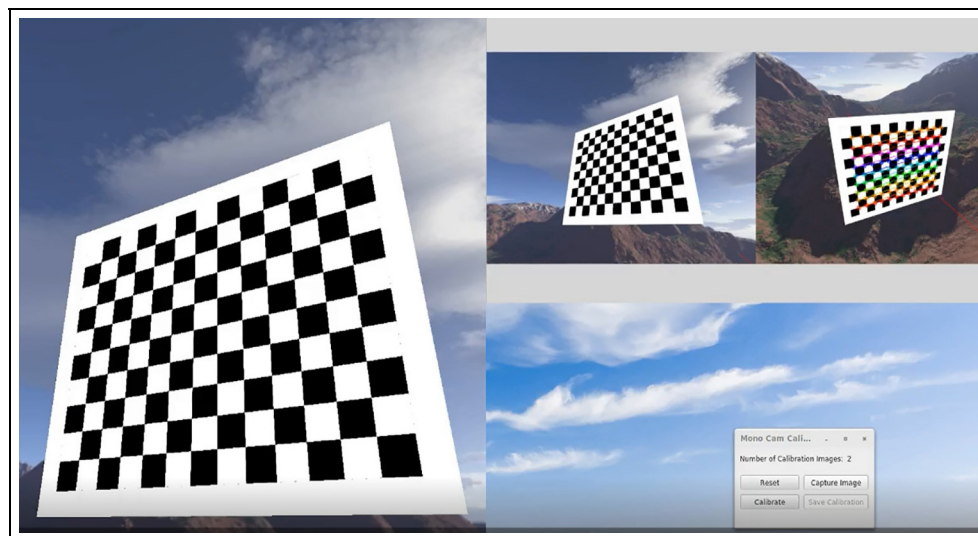


Figure 4. The camera calibration module. The left-hand window shows a camera in the virtual environment viewing a checkerboard. The right-hand windows show the calibration module.

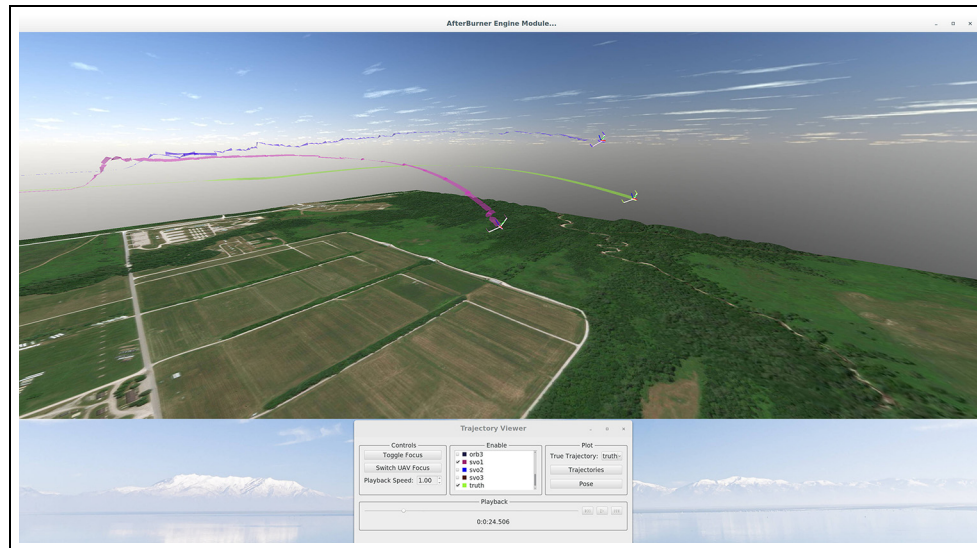


Figure 5. View of the simulated world with the trajectory viewer enabled. The ribbons outline the trajectories of the different visual odometry estimates.

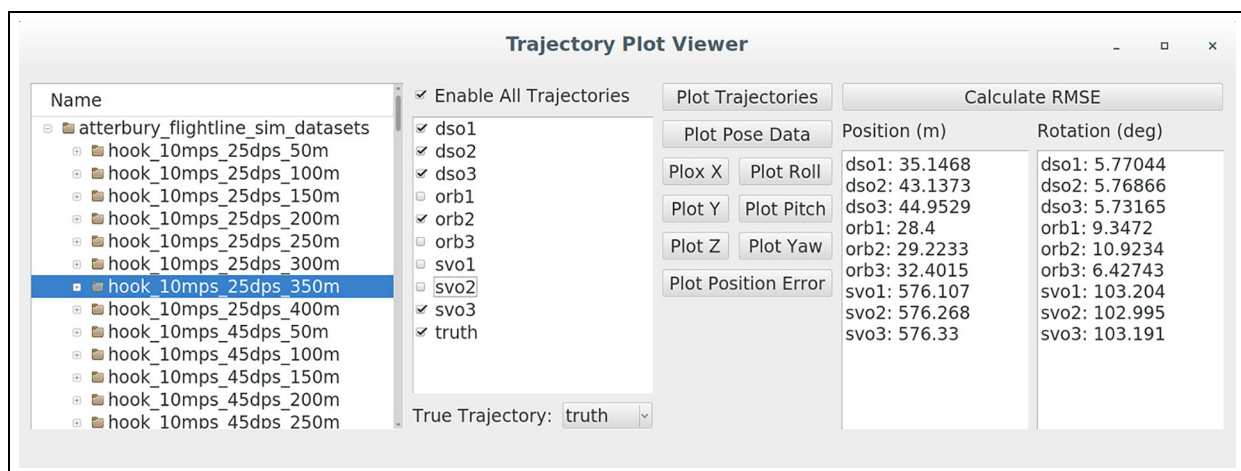


Figure 6. View of the trajectory viewer control panel.

panes contain controls for plotting and calculating the position and rotation errors for each of the trajectory files.

2.7 Visual odometry algorithms

The navigation testbed is configurable such that any VO algorithm that can take imagery via the Robot Operating System (ROS) or Lightweight Communications and Marshalling (LCM) should be pluggable within the system. Three different open-source, state-of-the-art algorithms are tested: DSO,¹ SVO,² and ORB-SLAM2.³ SVO is used instead of SVO2 since the source code for SVO2 is not currently open source. ORB-SLAM2 is allowed for

comparison against the other VO algorithms by disabling the loop closure thread so that it essentially functions as a feature-based VO algorithm.

Each of the VO algorithms are tested using their default parameter settings. Additional tests are conducted by adjusting individual parameters specific to each algorithm. In DSO, setting `_minGradHistAdd` controls the image gradient threshold for candidate point selection in an image. Tests are run using the default value of seven and a lowered value of five to allow the selection of points with a lower gradient, since many of the scenes viewed over Camp Atterbury's flightline have little texture.¹ The features from accelerated segment test (FAST) score

thresholds of ORB-SLAM2 and SVO are also tested at lower values for the same reason. ORBextractor.minThFAST in ORB-SLAM2¹⁵ and triangMinCornerScore in SVO are tested at their default values of seven and a lowered value of two.²

SVO controls the creation of new keyframes through the kfSelectMinDist parameter, which is the percentage of the average scene depth. A new frame is designated as a keyframe if it is farther than this distance from the previous keyframes. This is tested at its default value of 12% and a lowered value of 5% to create keyframes more frequently and increase the robustness of the algorithm under fast movements. Additional details for each algorithm are found below.

2.7.1 DSO. DSO was released by Engel et al.¹ in 2016. This algorithm not only utilizes geometric camera calibration to model the camera's intrinsic parameters, but also incorporates an optional photometric camera calibration to model the camera's exposure response and pixel attenuation (vignetting). Photometric camera calibration is not necessary in feature-based methods since features are selected to be invariant to changes in lighting conditions, but is helpful in direct methods since they depend solely on pixel intensities. In DSO, new frames are tracked relative to the latest keyframe. If the frame is not designated as a new keyframe, it is discarded. If a new keyframe is created, the photometric error is optimized over a sliding window of the latest seven keyframes. DSO tracks a fixed, sparse set of pixels across all keyframes. Candidate pixels in an image are selected for tracking by dividing the image into blocks of size $d \times d$ and calculating a region-adaptive gradient threshold for each block based on the median gradient of the pixels in the block. The pixel with the highest gradient greater than the threshold is chosen. To be able to include pixels in regions with small gradients, the process is repeated again over larger block sizes of $2d$ and $4d$ with weaker gradient thresholds. DSO is currently open source.

2.7.2 ORB-SLAM2. ORB-SLAM is one of the leading feature-based VSLAM algorithms and was created by Mur-Artal et al.¹⁵ in 2015. ORB-SLAM involves three threads for tracking, mapping, and loop closing. ORB-SLAM maintains an undirected weighted graph of keyframes that are linked to each other based on the number of shared map points in the keyframe images. This covisibility graph is used for loop closures and to pose graph optimizations. However, in order to increase efficiency, a subset of the covisibility graph is also maintained as a spanning tree and is called the essential graph. The essential graph contains all keyframe nodes but only connects keyframes sharing the most map points.

ORB-SLAM extracts Oriented FAST and Rotated BRIEF (binary robust independent elementary features) (ORB) features from an image and uses them for tracking, re-localization, and loop detection, which allows for high efficiency. In the tracking thread, an initial pose estimation is obtained for the current frame by either using a constant velocity motion model from the last frame if tracking was successful in the last frame or, if tracking was lost, by re-localizing from the bag-of-words recognition database. The pose estimation is optimized by projecting the local map of keyframes containing covisible map points into the current frame. The keyframe with the most covisible map points is designated as the reference keyframe. If the current frame's image is sufficiently different from that of the reference keyframe, it is inserted into the covisibility graph as a new keyframe in the mapping thread. Map points that have been determined to be non-trackable or erroneously triangulated are culled. Otherwise, new map points are triangulated from matching ORB features in connected keyframes. Local bundle adjustment is run to optimize the current keyframe along with the connected keyframes in the covisibility graph and all map points belonging to those keyframes. Redundant keyframes are then culled.

The loop closing thread utilizes the bag-of-words approach to query the recognition database, find similar images, and insert new edges into the covisibility graph for loop closures. The loop closure error is then propagated throughout the graph by optimizing the essential graph and transforming all map points according to its keyframe's correction. A new version of the algorithm, ORB-SLAM2, was released in 2017.³ ORB-SLAM2 adds in a full bundle adjustment step over all keyframes and map points in a separate fourth thread. It also extends the algorithm to accommodate monocular, stereo, and RGB-D cameras. ORB-SLAM and ORB-SLAM2 are both open source.

2.7.3 SVO. SVO is a hybrid VO algorithm by Forster et al.² that was first released in 2014. SVO consists of a motion estimation or tracking thread and a mapping thread. In the motion estimation thread, an initial estimate of the camera pose is found through sparse model-based image alignment in which a direct method is used to minimize the photometric error with reprojected feature patches from the previous image. The reprojected features in the new image are then aligned with respect to the rest of the map by optimizing the 2D pixel locations to minimize the photometric error with the reference feature patch in the keyframe. By adjusting each individual reprojected feature location in the new frame, the feature alignment step violates epipolar constraints. In order to correct this, SVO performs a bundle adjustment through the pose and structure refinement step, optimizing the camera pose again but this time by minimizing the photometric error with respect to optimized feature patches. If the mapping thread receives a

keyframe, it splits the image into fixed-size cells and extracts the FAST corners with the highest Shi-Tomasi score in each cell. Depth filters are initialized for new features with high uncertainty. New regular frames are used to update these depth filters using a Bayesian method and, once the variance is sufficiently low, the corresponding 3D point is inserted into the map to be used for motion estimation. By only extracting feature points during keyframes and using direct methods to calculate the camera pose for every frame, SVO is able to achieve high processing speeds. The authors also presented SVO 2.0 in 2017.¹⁸ This extends the original SVO algorithm to large FOV cameras, multi-camera systems, IMU incorporation, and the additional use of edges for feature alignment. Although the original SVO algorithm is open source, SVO 2.0 is only available as a pre-compiled binary.

2.8 Message protocols

A core issue for robotic systems is inter-process communication (IPC), since most systems require the development and use of multiple subsystems running as separate processes or nodes. Several open-source middleware programs have been developed to address this problem, and one of the most popular is the ROS.³⁷ IPC in the ROS is accomplished through a message-passing system where language-agnostic message types can be built and used to generate the appropriate language-specific data structures and files to be imported by nodes for communication. Messages are passed from publisher to subscriber nodes that are on the same topic. Key design features in the ROS are that all communications between nodes are managed by a centralized node and that the principal protocol used is the Transmission Control Protocol (TCP). This makes the ROS less than ideal for building distributed systems with potentially unreliable connections between nodes. However, the ROS features an extensive mature library beyond IPC with a large open-source community, provides a centralized parameter server and launch system for easy management and distribution of data among nodes, and maintains a build system augmenting CMake for support in importing ROS packages and libraries.

Another open-source middleware is the LCM library.³⁸ This was designed by students at the Massachusetts Institute of Technology (MIT) to offer the bare essentials of just IPC. LCM operates using a similar system as the ROS, passing messages from publisher nodes to subscriber nodes that are on the same channel. Communication between nodes is accomplished through the User Datagram Protocol (UDP) in a decentralized network, making LCM more suitable for distributed systems. However, LCM lacks the presence of the ROS's large open-source community and libraries. LCM also lacks a

centralized parameter server, requiring alternative methods to distribute parameters to disparate nodes.

The lines depicted in Figure 1 represent message passing through one of these two transport protocols. The ROS is used for its launch system and parameter server to be able to correctly initialize and launch multiple nodes. It is also used to take advantage of the MAVROS library, an extensive and mature library for communicating with the ArduPilot autopilot and simulator. LCM is used for compatible communication with systems being developed at the AFIT ANT Center. ROS Lunar and LCM version 1.3.1 are used in this work.

2.9 Implementation details

All modules were built and tested on a laptop computer with an Intel Core i7-7500U processor and a Nvidia Quadro M2200 GPU. The operating system used was Ubuntu 16.04. All modules were programmed in C++. IPC between the individual nodes was accomplished mainly through the ROS.³⁷ However, IPC with the real-world autopilot, sensors, and camera was accomplished through the LCM library.³⁸

Simulated trajectories are generated by interfacing the virtual environment with the MAVROS library and Ardupilot autopilot software. Ardupilot provides a SITL module that simulates an autopilot. SITL can be programmed with missions and controlled through a standard GCS to conduct simulated flight tests with realistic flight dynamics. Figure 2 shows an example in which a GCS is used to plan a flight and control the aircraft. The left-hand window shows the virtual environment, the top right-hand window shows the live image feed from the virtual camera attached to the bottom of the aircraft, and the bottom right-hand window shows QGroundControl, the GCS.

3 Experiments

Flight tests were conducted both in simulation and in the real world over Camp Atterbury, Indiana, with a fixed-wing aircraft and a camera facing straight downward relative to the aircraft's body frame. The objectives of the flight tests are two-fold: to demonstrate the utility of the virtual world flights in evaluating visual navigation approaches and to identify the best state-of-the-art VO algorithm for use on a fixed-wing UAV. These objectives were evaluated by analyzing the real-time performance of each algorithm under high speeds and altitudes while performing aggressive maneuvers.

3.1 Assumptions

A consideration that has to be taken into account is the inability of monocular visual navigation algorithms to be able to determine absolute scale without external input. All

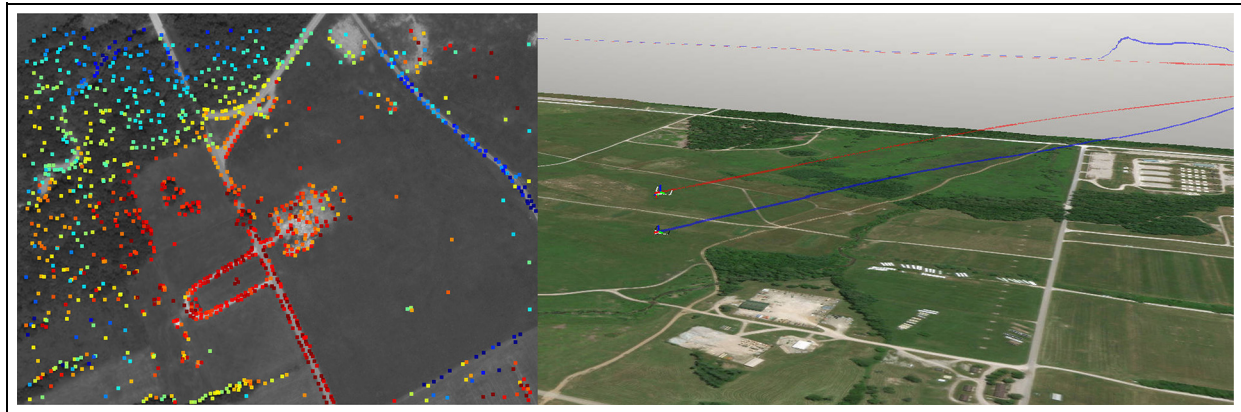


Figure 7. Simulated flight with Direct Sparse Odometry (DSO). The left-hand image is an output of DSO, showing the points of interest for the navigation. The right-hand image shows the virtual world and the comparison between truth and estimates. The red small unmanned aircraft system ribbon shows the truth trajectory and the blue shows the estimate from DSO. (Color online only.)

three algorithms compute trajectory positions to a relative scale. To account for this, all flight tests included an initialization leg in which the aircraft flies straight and level such that the camera points straight down toward Earth. The VO algorithms are initialized during this leg and the altitude at which the algorithm is initialized is used as the absolute scale for the entire trajectory. Furthermore, the loop closure capability in ORB-SLAM2 was disabled in order to make a fair comparison against the other strict VO algorithms.

Both SVO and DSO were able to process images and produce state estimates at the frame rate of the input imagery stream, 31 fps. However, ORB-SLAM2 was limited to producing state estimates at a rate of 20 fps.

A small search over several different settings and parameters for each VO algorithm was conducted.³⁹ The default parameters for each of the algorithms offered the best performance in both the simulation and real-world flight tests and all reported results found below were obtained from using the default settings.

3.2 Simulation setup

Figures 7–9 show examples of running each of the algorithms on simulated flight trajectories and reprojecting the pose estimates back into the virtual environment. The left-hand windows show the processed images from the VO algorithms with the tracked pixels or features in the current image. The right-hand windows show the virtual environment with the truth trajectory represented by the SUAS with the red trailing ribbon and the estimated trajectory by the SUAS with the blue trailing ribbon.

The virtual camera settings used are shown in Table 1. The virtual environment was able to continuously supply a framerate of 31 Hz of simulated imagery during all the simulation experiments discussed below.

The algorithms were tested on a 3.3 km trajectory, shown in Figure 10, where the SUAS starts at the north end of the runway, flies south for the initialization leg, and then returns north in an oscillating pattern. The experimental parameters that were varied are shown in Table 2. In this simulation, we did not add wind and the trajectory is relatively flat, so airspeed is equal to ground speed. The roll rate setting limits the maximum roll rate allowed by the SITL autopilot.

3.3 Real-world flight test

The VO algorithms were also tested on data gathered from real-world flight tests on a 1/3-scale Carbon Cub aircraft weighing 50 lbs with a wingspan of 14 ft. A Pixhawk 2 autopilot was used for low-level control of the aircraft and all payload data was stored in LCM log files for post-flight playback and analysis. The LCM log files were recorded in-flight on an onboard Intel NUC7 computer with an i7 processor. Grayscale imagery was collected from a Prosilica GT1290 camera whose properties are listed in Table 3. Truth data was provided by a Piksi navigation board with GPS data at 10 Hz and IMU data at 100 Hz. All data was time-stamped using a TM2000A Precision Time Protocol (PTP) server. Processing of the flight data was conducted on the laptop computer.

The algorithms were tested on a 9.6 km trajectory, shown in Figure 11, where the aircraft starts at the western end of the airfield and flies three counter-clockwise loops while moving east. The trajectory was flown at 250 m in altitude with airspeeds varying between 20 and 30 m/s for a flight time of 284 seconds.

The weather the day of the flight testing provided additional challenges. Much of the imagery from the few successful flights is close to saturation due the exposure time

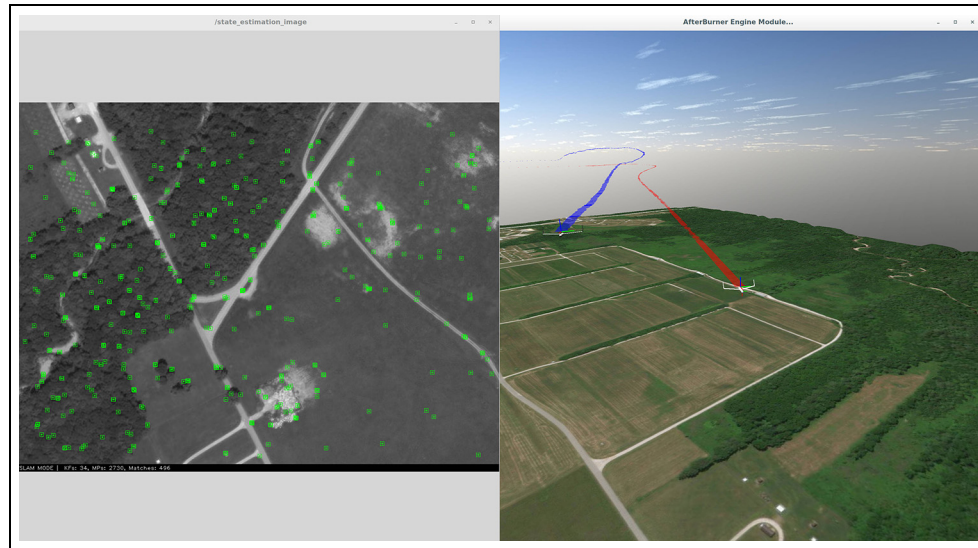


Figure 8. Simulated flight with ORB-SLAM2. The left-hand image is an output of Direct Sparse Odometry, showing the points of interest for the navigation. The right-hand image shows the virtual world and the comparison between truth and estimates. The red small unmanned aircraft system ribbon shows the truth trajectory and the blue shows the estimate from ORB-SLAM2. (Color online only.)

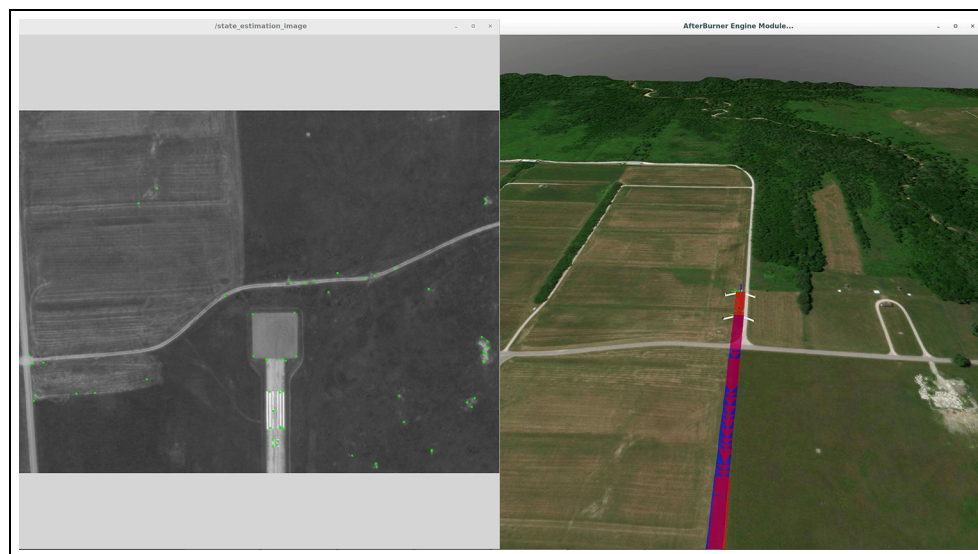


Figure 9. Simulated flight with Semi-Direct Visual Odometry (SVO). The left-hand image is an output of Direct Sparse Odometry, showing the points of interest for the navigation. The right-hand image shows the virtual world and the comparison between truth and estimates. The red small unmanned aircraft system ribbon shows the truth trajectory and the blue shows the estimate from SVO. (Color online only.)

set on the camera to achieve its maximum frame rate of 33 fps. This leads to a saturated image on bright days that, when combined with the low texture environment of the airfield, provides a difficult challenge for the VO algorithms to be able to detect and track features or pixels. We did not attempt to recreate the saturated images from the flight test

with the simulation environment, as one of the main objectives of the study was to evaluate VO performance. However, the capabilities of the simulation would have allowed for that possibility. Future work will examine simulating different sensor failures to evaluate for robustness of algorithms when faced with challenging sensor information.

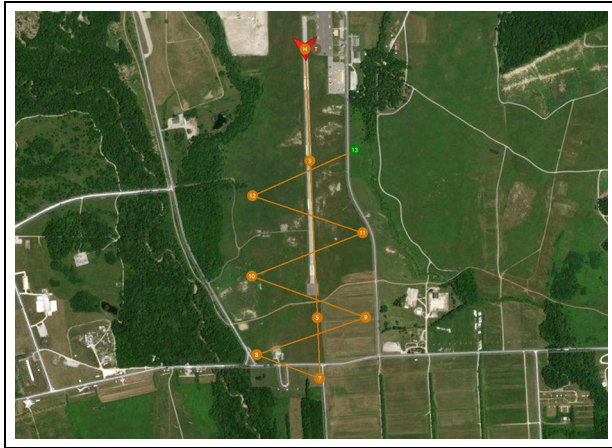


Figure 10. One sample of a simulation trajectory use during this study. The results shown are from this trajectory. Additional trajectories examined are found in the associated thesis for this work.³⁹

Table 1. Virtual camera settings.

Option	Setting
Resolution	1280 × 960
Aspect ratio (width/height)	1.333
Field of view (deg)	84.872
Frame rate (fps)	31

Table 2. Simulation experimental parameters.

Experimental parameters	Settings
Airspeed (m/s)	10, 25
Maximum roll rate (deg/s)	25, 45, 65
Altitude	50–400 m in 50 m intervals

4 Experimental results

This section discusses the simulation results first and then the real-world flight results. Algorithm performance is reported through the 3D root-mean-square error (RMSE) of the position in meters and the rotation in degrees from the ground truth throughout the entire trajectory. Overall, ORB-SLAM2 proved to be the most flexible and robust algorithm throughout a variety of conditions in both simulation and real-world flight tests.

4.1 Simulation

The results for the simulation flights with a slower airspeed of 10 m/s are shown in Table 4. If an algorithm

Table 3. Real-world camera properties.

Property	Setting
Resolution	1280 × 960
Aspect ratio (width/height)	1.333
Lens field of view (deg)	84.872
Lens sensor format (in)	½
Camera sensor format (in)	1/3
Frame rate (fps)	33

failed to initialize and track throughout the entire trajectory, the position and rotation errors are marked with an X. In addition, the entries for the algorithm with the lowest position and rotation errors for a trajectory are highlighted. Most of the algorithms encountered difficulties initializing at lower altitudes due to the limited resolution of the satellite imagery and the limited FOV.

SVO had the most difficulty and was unable to initialize below 300 m in altitude. SVO also proved to be the most fragile algorithm and experienced great difficulty maintaining localization, especially during rotations. This is due to the method of keyframe selection. SVO creates new keyframes strictly based on the Euclidean distance between the newest frame and the previous keyframes, failing to take rotations into account. For this reason, SVO was never able to track past the first turn.

DSO and ORB-SLAM2 were able to successfully track the entire trajectory for the majority of the test cases. ORB-SLAM2 provided the most accurate position estimates throughout all trajectories in which it was able to initialize. Although ORB-SLAM2 and DSO produced comparable rotation estimates, with ORB-SLAM2 being slightly more accurate at lower roll rates, DSO provided a more consistent and accurate estimate at the highest roll rate of 65 deg/s.

Figures 12 and 13 show a representative trajectory from this dataset along with its individual pose data to illustrate the VO performances in detail. This data is obtained from the SUAS flying at 300 m in altitude with a maximum roll rate of 25 deg/s. SVO fails 23 seconds into the trajectory before entering the first turn. DSO drifts further from the truth's position than ORB-SLAM2 as the trajectory progresses, while both DSO and ORB-SLAM2 maintain close approximations of the rotation.

Table 5 shows the VO performance results at a higher airspeed of 25 m/s. The operational limit at which the algorithms were able to maintain tracking throughout the trajectories was at the maximum roll rate of 45 deg/s for this speed. ORB-SLAM2 was able to maintain localization throughout most of the trajectories at this roll rate, while DSO experienced frequent failures and SVO continued to fail on the first turns. All algorithms had difficulty



Figure 11. Path flown for the real-world results.

Table 4. Simulation visual odometry (VO) results at 10 m/s. The position and rotation errors for trajectories in which a VO algorithm failed to initialize are marked with an X. The algorithm with the lowest position or rotation error for a trajectory is highlighted.

Roll rate (deg/s)	Altitude(m)	Position RMSE (m)			Rotation RMSE (deg)		
		DSO	ORB	SVO	DSO	ORB	SVO
25	50	X	X	X	X	X	X
	100	170.997	X	X	4.075	X	X
	150	X	60.898	X	X	4.914	X
	200	113.955	28.420	X	5.541	64.059	X
	250	110.229	20.434	X	6.504	5.470	X
	300	127.009	20.137	403.935	7.965	7.113	188.684
	350	126.387	18.376	545.184	8.274	5.717	187.95
	400	132.861	15.331	539.435	12.257	6.784	182.351
45	50	X	X	X	X	X	X
	100	152.962	X	X	5.101	X	X
	150	104.605	93.130	X	26.227	28.169	X
	200	136.505	28.474	X	10.358	5.355	X
	250	106.315	14.406	X	8.429	6.415	X
	300	146.665	15.811	428.557	11.835	5.923	184.938
	350	119.18	17.349	520.012	7.558	22.937	198.616
	400	176.371	15.749	535.812	17.827	6.922	183.166
65	50	X	X	X	X	X	X
	100	182.635	X	X	83.158	X	X
	150	124.558	77.100	X	27.253	84.419	X
	200	131.615	15.688	X	6.216	6.455	X
	250	83.809	18.574	X	25.237	70.235	X
	300	138.955	15.458	527.49	11.463	6.142	186.732
	350	133.632	19.362	514.676	18.412	72.723	185.18
	400	154.969	16.639	516.447	23.975	26.571	192.534

RMSE: root-mean-square error; DSO: Direct Sparse Odometry; ORB: ORB-SLAM2; SVO: Semi-Direct Visual Odometry.

maintaining localization at roll rates of 65 deg/s. ORB-SLAM2 continued to maintain a more accurate position estimate than DSO. DSO exhibited a slightly more

accurate rotation estimate at lower roll rates of 25 deg/s, although this was reversed at 45 deg/s with ORB-SLAM2 providing significantly better estimates.

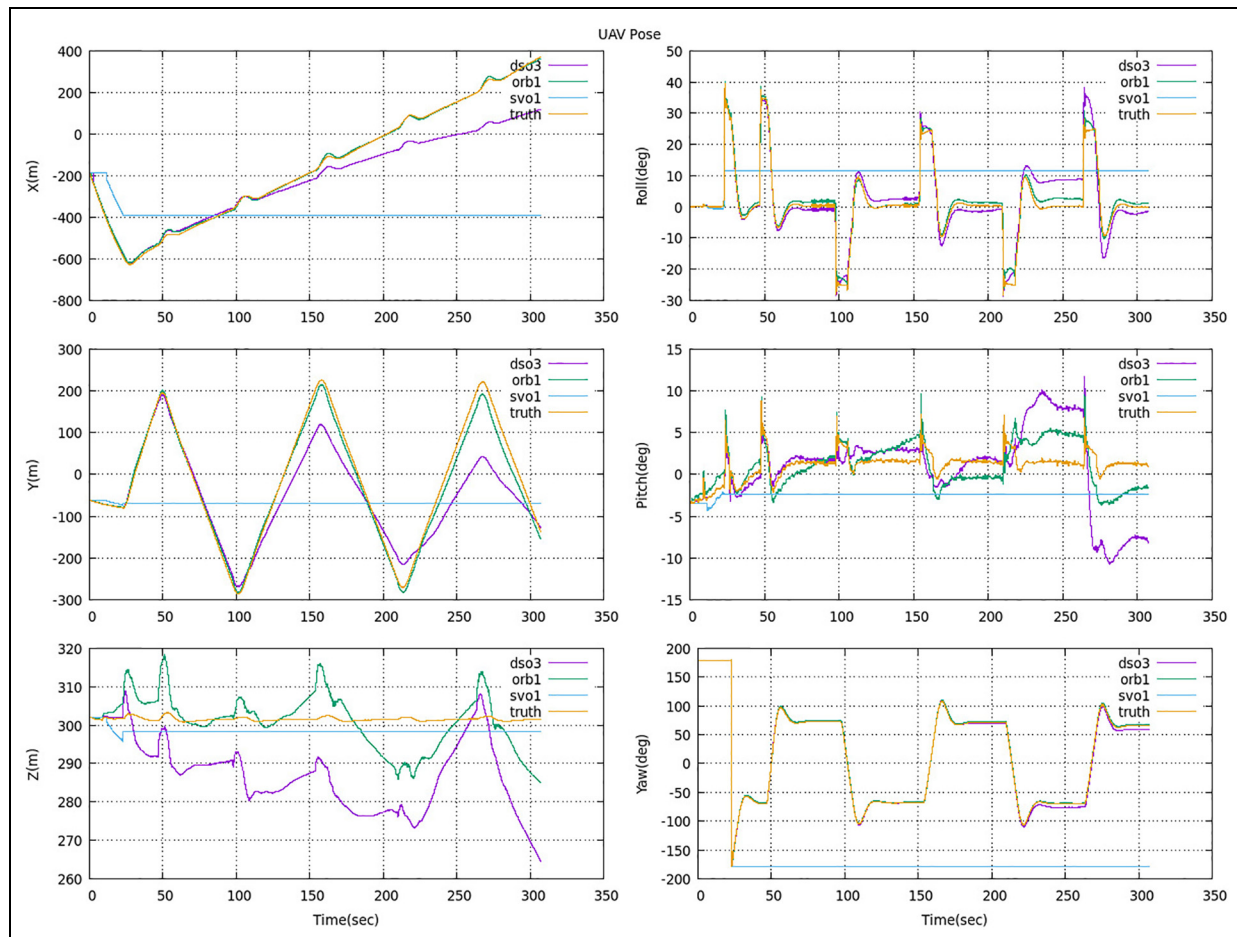


Figure 12. Simulation trajectories at 10 m/s, 25 deg/s, 300 m for Direct Sparse Odometry, Semi-Direct Visual Odometry, and ORB-SLAM2 compared to truth.

The simulation experiments clearly identified that ORB-SLAM2 provided the most robust capability to adapting to the new domain of fixed-wing flight. However, the simulation framework also allowed for accomplishing more than identifying the best candidate VO algorithm for use in fixed-wing environments. Hundreds of runs in different configurations and slightly different trajectories were run within days. A comparable study in real life, using live camera data from an aircraft, would require a tremendous amount of resources: it would be cost prohibited. With the virtual world we were also able to hold many conditions constant, such as lighting conditions, timing, and the actual trajectory. As will be discussed further below, this is a key feature for the simulation.

4.2 Real-world flight test

All three algorithms experienced difficulties initializing and tracking throughout the gathered real-world dataset

due to the previously discussed problem of image saturation. DSO failed to initialize throughout any of the real-world imagery. SVO was able to initialize but failed to maintain tracking for more than a few seconds. ORB-SLAM2 was the most successful and was able to track through most of the trajectory, but still did not manage to get an ideal initialization. Figure 14 illustrates ORB-SLAM2's estimated trajectory against the truth trajectory data. Figure 15 shows the detailed pose estimates. The most extreme errors occur in the z-axis since the algorithm initially miscalculated the pitch, leading to oscillations in altitude between the north and south ends of the trajectory. The total position RMSE was 321 m and the rotation RMSE was 112 deg for ORB-SLAM2.

For comparison, ORB-SLAM2 was also run on virtual imagery from the SUAS flying the real-world trajectory dataset in the virtual environment. Both trajectories estimated from the virtual and real-world imagery are shown in Figure 16 and their pose data is shown in Figure 17.

Table 5. Simulation visual odometry (VO) performance at 25 m/s. The position and rotation errors for trajectories in which a VO algorithm failed to initialize are marked with an X. The algorithm with the lowest position or rotation error for a trajectory is highlighted.

Roll rate (deg/s)	Altitude (m)	Position RMSE (m)			Rotation RMSE (m)		
		DSO	ORB	SVO	DSO	ORB	SVO
25	50	X	X	X	X		X
	100	X	X	X	X	X	X
	150	X	81.206	X	X	7.047	X
	200	167.809	71.906	X	4.535	8.326	X
	250	176.991	65.365	X	23.527	31.138	X
	300	177.575	62.717	368.72	7.012	9.573	181.637
	350	167.853	58.907	509.857	7.111	48.707	182.546
	400	173.911	40.369	514.787	7.047	13.425	182.213
45	50	X	X	X	X	X	X
	100	212.844	X	X	82.186	X	X
	150	X	210.371	X	X	79.259	X
	200	X	100.056	X	X	49.045	X
	250	381.837	90.671	X	83.847	48.052	X
	300	407.325	61.127	386.2	87.066	11.731	182
	350	196.225	47.466	505.137	9.248	7.815	180.623
	400	313.935	45.072	497.773	78.373	41.819	189.143
65	50	X	X	X	X	X	X
	100	235.459	X	X	59.036	X	X
	150	X	419.035	X	X	129.506	X
	200	444.963	443.574	X	88.568	185.526	X
	250	482.253	519.21	X	90.544	188.444	X
	300	390.829	426.205	506.941	86.453	118.167	187.349
	350	513.123	194.008	506.12	181.218	89.653	182.838
	400	499.621	508.358	498.452	179.734	186.319	180.706

RMSE: root-mean-square error; DSO: Direct Sparse Odometry; ORB: ORB-SLAM2; SVO: Semi-Direct Visual Odometry.

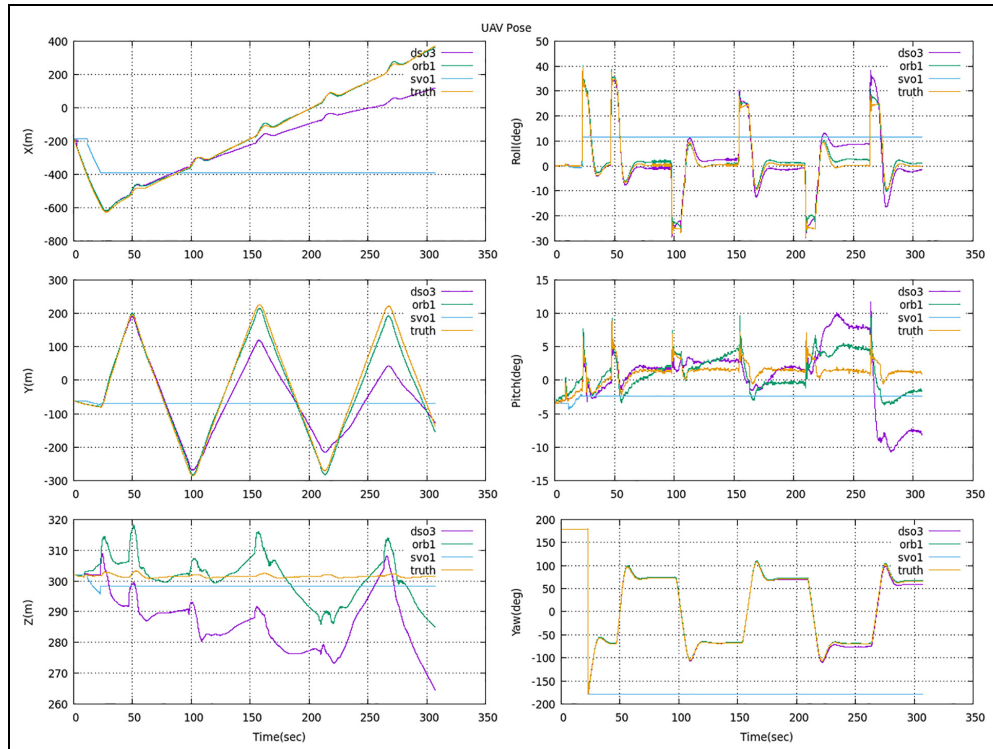


Figure 13. Simulation pose data at 10 m/s, 25 deg/s, 300 m. Note that the scales are different for each plot. For example, the scale of the z position plot is much smaller than that of x or y, greatly emphasizing the difference in values.

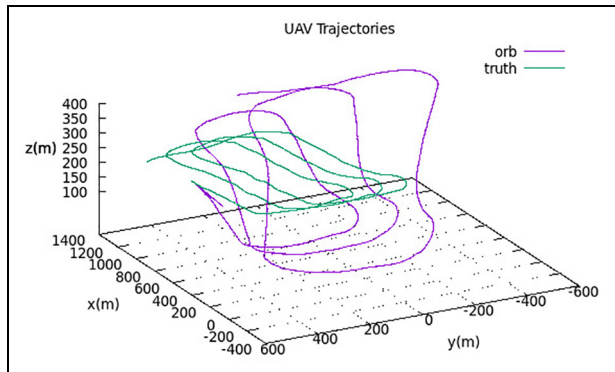


Figure 14. Real-world ORB-SLAM2 trajectory compared to truth.

The algorithm was able to obtain better initialization using the virtual imagery since the lighting conditions were constant and it was easier to detect features. However, it lost

localization after the second loop due to erratic jumps, likely due to the combination of vibrations and sensor noise in the trajectory data. To draw better comparisons of algorithm performance between the virtual and real world, ORB-SLAM2 needs to be run on trajectory data that is smoothed and gathered at higher frequencies along with imagery corrected for saturation.

4.3 Analysis

Overall, ORB-SLAM2 proves to be the most flexible VO algorithm that is suitable for fixed-wing applications. Both simulation and real-world flight tests show a robust initializer and tracking capability. Although it is not able to achieve the frame rates of SVO or DSO, ORB-SLAM2 is able to track far more robustly at higher speeds and sharper turns, where the combination of 25 m/s and 45 deg/s is the operational limit. Its robustness is further highlighted by its capability to maintain localization under a greatly throttled image stream

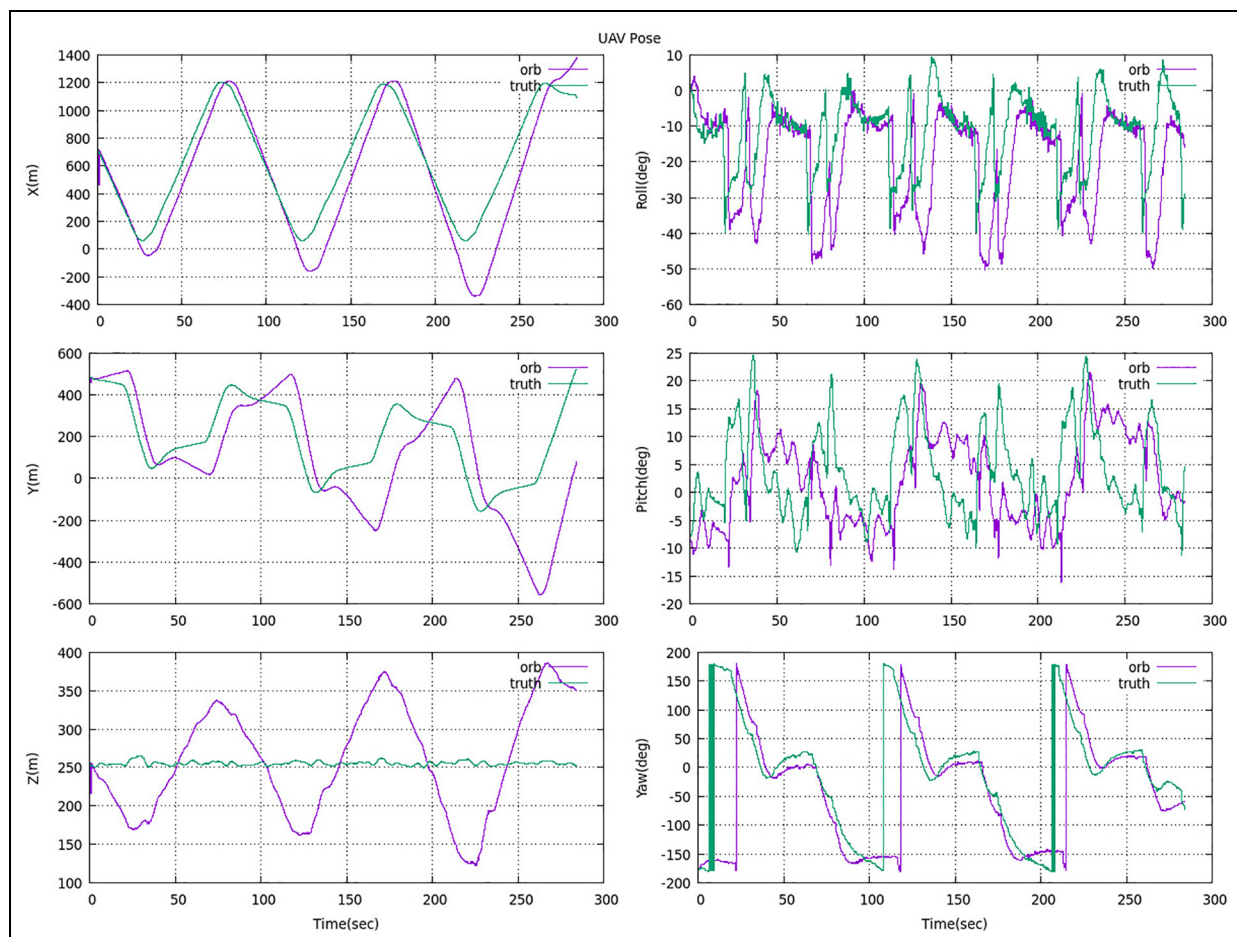


Figure 15. Real-world ORB-SLAM2 pose data results, compared to truth. The most extreme errors occur in the z-axis, since the algorithm initially miscalculated the pitch, leading to oscillations in altitude between the north and south ends of the trajectory.

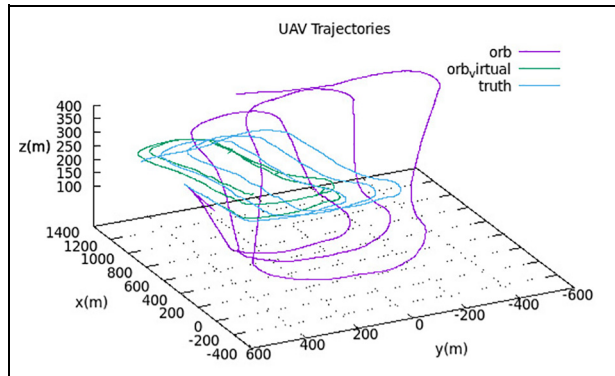


Figure 16. ORB-SLAM2 trajectory from real-world trajectory and virtual imagery.

rate. ORB-SLAM2 also has the distinct advantage in further improvement, since a SLAM scheme is incorporated naturally in the design of the algorithm.

SVO is fast but proves to be very fragile, especially during rotations. SVO is unable to cope with any of the turns during simulation or real-world flight tests, even at the slowest speeds and roll rates. This is primarily due to its method of designating new keyframes by a measure of distance. If using SVO for fixed-wing applications, this heuristic must be modified to another method, such as that of ORB-SLAM2 or DSO, where keyframe creation is based on the measure of changes in its FOV.

DSO's greatest weakness lies in the initializer. The authors' note on their git repository² is that the current initializer is not very good and requires slow, easy movements with a great deal of translation and little rotation. This is much easier to achieve in simulation with perfectly

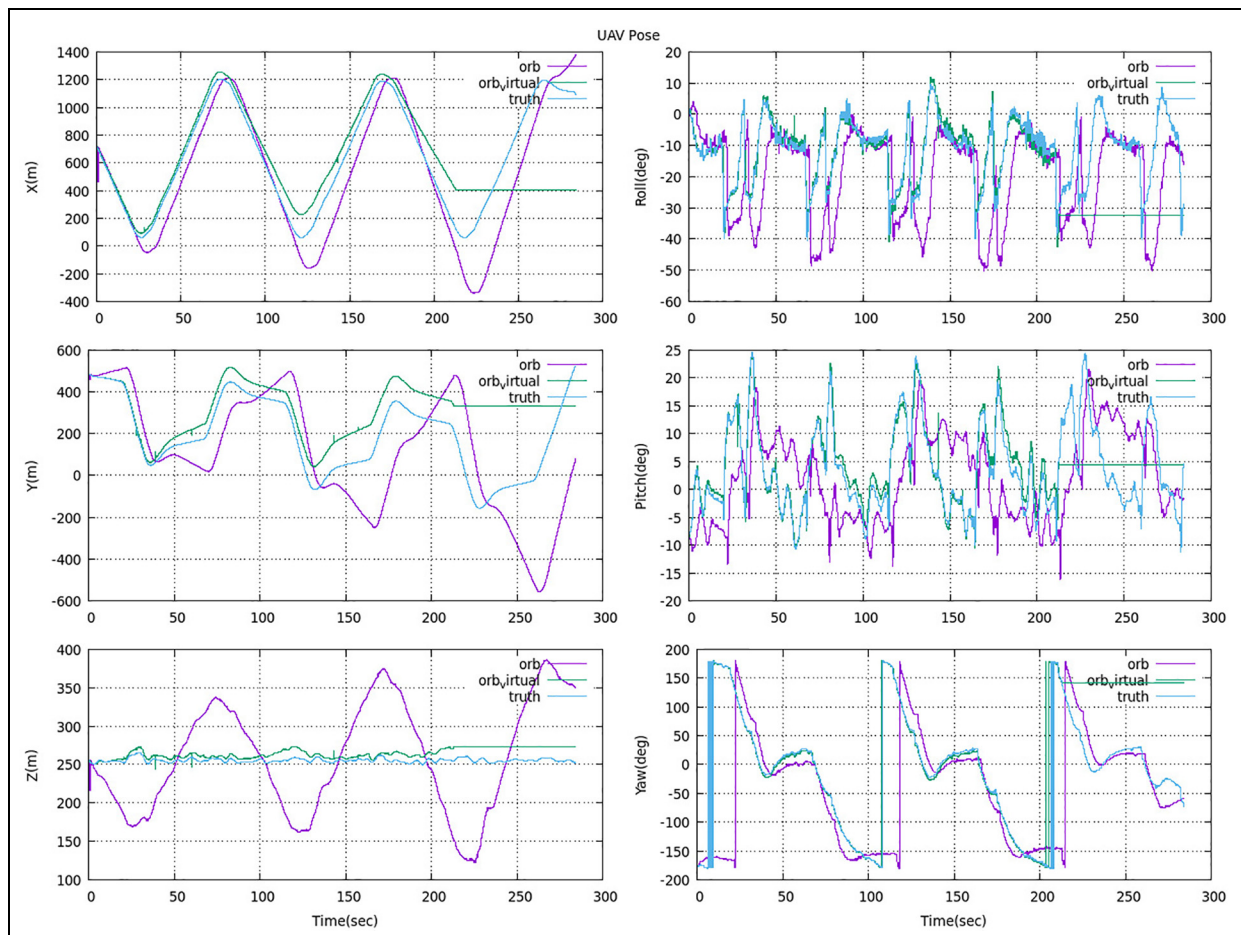


Figure 17. ORB-SLAM2 pose data from real-world trajectory and virtual imagery. The algorithm was able to obtain better initialization using the virtual imagery since the lighting conditions were constant and it was easier to detect features. However, it lost localization after the second loop due to erratic jumps, likely due to the combination of vibrations and sensor noise in the trajectory data.

smooth flights creating ideal environments for initialization. However, real-world flights prove to be much more difficult as the SUAS invariably encounters turbulence from wind. This is shown by DSO being unable to successfully initialize in any of the real-world datasets. If using DSO, the initializer should be replaced by a faster method, such as through feature-based homography or random sample consensus (RANSAC). The initializer could also be either augmented or completely replaced by IMU measurements. Another drawback to DSO is that it is not as amenable to incorporating loop closures as other feature-based VO algorithms, since many of the current loop closure frameworks make use of features for the bag-of-words approach.⁴⁰


5 Conclusion


In this paper, we presented a virtual environment for real-time testing and analysis of VO and VSLAM algorithms. This was used to test the performance of three open-source algorithms: DSO, SVO, and ORB-SLAM2 (with loop closures disabled). The algorithms were tested under various airspeeds, roll rates, and altitudes. The algorithms were also tested on a real-world dataset and the results compared against virtual data. Through these tests, ORB-SLAM2 was found to be the most flexible and robust algorithm in both simulation and real-world flights. Although this work demonstrated the flexibility of the virtual testbed, further work must be accomplished to draw better correlations and conclusions between the state estimates obtained from flight tests in the virtual environment and the real world, primarily obtaining a filtered real-world dataset with a more accurate and smoothed ground truth along with imagery corrected for saturation.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

ORCID iDs

Robert C Leishman  <https://orcid.org/0000-0002-5363-1725>

Scott L Nykl  <https://orcid.org/0000-0002-4271-2132>

References

- Engel J, Koltun V and Cremers D. Direct sparse odometry. *IEEE Trans Pattern Anal Mach Intell* 2018; 40: 611–625.
- Forster C, Pizzoli M and Scaramuzza D. SVO: fast semi-direct monocular visual odometry. In: *proceedings of the IEEE international conference on robotics and automation*, Hong Kong, China, September 2014, pp.15–22. IEEE.
- Mur-Artal R and Tardos JD. ORB-SLAM2: an open-source SLAM system for monocular, stereo, and RGB-D cameras. *IEEE Trans Robot* 2017; 33: 1255–1262.
- Hartley R and Zisserman R. *Multiple view geometry in computer vision*. 2nd ed. Cambridge: Cambridge University Press, 2004.
- Scaramuzza D and Fraundorfer F. Visual odometry. *IEEE Robot Autom Mag* 2011; 13: 80–92.
- Scaramuzza D and Fraundorfer F. Visual odometry part II. *IEEE Robot Autom Mag* 2011; 18: 80–92.
- Durrant-Whyte H and Bailey T. Simultaneous localization and mapping ({SLAM}): part {I}. *Robot Autom Mag* 2006; 13: 99–110.
- Bailey T and Durrant-Whyte H. Simultaneous localization and mapping ({SLAM}): part {II}. *IEEE Robot Autom Mag* 2006; 13: 108–117.
- Hassaballah M, Abdelmgeid AA and Alshazly HA. *Image feature detectors and descriptors*. Volume 630. 2016. Switzerland: Springer.
- Irani M and Anandan P. All about direct methods. In: *ICCV workshop on vision algorithms*, Corfu, Greece, 21–22 September 1999, pp.267–277. Berlin: Springer-Verlag.
- Klein G and Murray D. Parallel tracking and mapping for small AR workspaces. In: *proceedings of the 2007 6th IEEE and ACM international symposium on mixed and augmented reality*, 13–16 November 2007. Nara, Japan: IEEE.
- Newcombe RA, Lovegrove SJ and Davison AJ. DTAM: dense tracking and mapping in real-time. In: 2011 international conference on computer vision, Barcelona, Spain, 6–13 November 2011, pp.2320–2327. IEEE. DOI: 10.1109/ICCV.2011.6126513.
- Engel J, Schops T and Cremers D. LSD-SLAM: Large-Scale Direct Monocular SLAM. In: *13th, European conference on computer vision – ECCV 2014*, Zurich, Switzerland, 2014, pp.834–849. Cham: Springer International Publishing.
- Engel J, Usenko V and Cremers D. A photometrically calibrated benchmark for monocular visual odometry. *Computing Research Repository (CoRR)* 2016. DOI:10.1016/j.aqpro.2013.07.003 .
- Mur-Artal R, Montiel JM and Tardos JD. ORB-SLAM: a versatile and accurate monocular SLAM system. *IEEE Trans Robot* 2015; 31: 1147–1163.
- Pizzoli M, Forster C and Scaramuzza D. REMODE: probabilistic, monocular dense reconstruction in real time. In: 2014 IEEE international conference on robotics and automation (ICRA), Hong Kong, China, 31 May–7 June 2014, pp. 2609–2616. IEEE. DOI: 10.1109/ICRA.2014.6907233.
- Faessler M, Fontana F, Forster C, et al. Autonomous, vision-based flight and live dense 3D mapping with a quadrotor micro aerial vehicle. *J Field Robot* 2015; 23: 1–20.
- Forster C, Zhang Z, Gassner M, et al. Semi-direct visual odometry for monocular, wide-angle, and multi-camera systems. *IEEE Trans Robot* 2017; 33: 249–265.
- Wheeler DO and Koch DP. Relative navigation of autonomous GPS-degraded micro air vehicles. *IEEE Contr Syst Mag* 2018; 38: 30–48.
- Ellingson G, Brink K and McLain T. Relative visual-inertial odometry for fixed-wing aircraft in GPS-denied

- environments. In: *2018 IEEE/ION position, location and navigation symposium (PLANS)*, Monterey, CA, 23–26 April 2018, pp.786–792. IEEE. DOI: 10.1109/PLANS.2018.8373454.
21. Wang S, Clark R, Wen H, et al. DeepVO: towards end-to-end visual odometry with deep Recurrent Convolutional Neural Networks. In: *2017 IEEE international conference on robotics and automation (ICRA)*, Singapore, 29 May–3 June 2017, pp.2043–2050. IEEE. DOI: 10.1109/ICRA.2017.7989236.
 22. Geiger A, Lenz P, Stiller C, et al. Vision meets robotics: the KITTI dataset. *Int J Robot Res* 2013; 32: 1231–1237.
 23. Sturm J, Engelhard N, Endres F, et al. A benchmark for the evaluation of RGB-D SLAM systems. In: *2012 IEEE/RSJ international conference on intelligent robots and systems*, Vilamoura, Portugal, 7–12 October 2012, pp.573–580. IEEE. DOI: 10.1109/IROS.2012.6385773.
 24. Burri M, Nikolic J, Gohl P, et al. The EuRoC micro aerial vehicle datasets. *Int J Robot Res* 2016; 35: 1157–1163.
 25. Majdik A, Till C and Scaramuzza D. The Zurich urban micro aerial vehicle dataset. *I J Robot Res* 2017; 36: 269–273.
 26. Gaidon A, Wang Q, Cabon Y, et al. Virtual Worlds as proxy for multi-object tracking analysis. In: *2016 IEEE conference on computer vision and pattern recognition (CVPR)*, Las Vegas, NV, 27–30 June 2016, pp.4340–4349. IEEE. DOI: 10.1109/CVPR.2016.470.
 27. Sayre-McCord T, Guerra W, Antonini A, et al. Visual-inertial navigation algorithm development using photorealistic camera simulation in the loop. In: *2018 IEEE international conference on robotics and automation (ICRA)*, Brisbane, QLD, Australia, 21–25 May 2018, pp.2566–2573. IEEE. DOI: 10.1109/ICRA.2018.8460692.
 28. Shah S, Dey D, Lovett C, et al. AirSim: high-fidelity visual and physical simulation for autonomous vehicles. In: Marco Hutter and Roland Siegwart (eds) *Field and service robotics*. Cham: Springer International Publishing, 2018, pp.621–635.
 29. Zhang Z, Rebecq H, Forster C, et al. Benefit of large field-of-view cameras for visual odometry. In: *2016 IEEE international conference on robotics and automation (ICRA)*, Stockholm, Sweden, 16–21 May 2016, pp.801–808. IEEE. DOI: 10.1109/ICRA.2016.7487210.
 30. Koenig N and Howard A. Design and use paradigms for Gazebo, an open-source multi-robot simulator. In: *2004 IEEE/RSJ international conference on intelligent robots and systems (IROS) (IEEE Cat. No.04CH37566)*, Sendai, Japan, 28 September–2 October 2004, Vol. 3, pp.2149–2154. IEEE. DOI: 10.1109/IROS.2004.1389727.
 31. Furrer F, Burri M, Achtelik MW, et al. RotorS – a modular Gazebo MAV simulator framework. In: Koubaa A (ed.) *Robot Operating System (ROS): the complete reference (Volume 1)*. Cham: Springer International Publishing, 2016, pp.595–625.
 32. Nykl S, Mourning C, Leitch M, et al. An overview of the STEAMiE educational game engine. In: *2008 38th annual frontiers in education conference*, Saratoga Springs, NY, November 2008, pp.F3B-21F3B-25. DOI: 10.1109/FIE.2008.4720454.
 33. Johnson DT, Nykl SL and Raquet JF. Combining stereo vision and inertial navigation for automated aerial refueling. *J Guid Contr Dynam* 2017; 40: 2250–2259.
 34. Paulson Z, Nykl S, John P, et al. Mitigating the effects of boom occlusion on automated aerial refueling through shadow volumes. *J Defens Model Simulat* 2018; 0: 1–15.
 35. Roeber J, Nykl S and Graham FS. Assessment of structure from motion for reconnaissance augmentation and bandwidth usage reduction. *J Defens Model Simulat* 2019; 1: 1–13.
 36. Parsons C, Paulson Z, Nykl SL, et al. Analysis of simulated imagery for real-time vision-based automated aerial refueling. *J Aerosp Inf Syst* 2019; 16: 77–93.
 37. Quigley M, Conley K, Gerkey B, et al. ROS: an open-source Robot Operating System. In: *ICRA workshop on open source software*, Kobe, Japan, 12 May 2009, vol. 3, p.5.
 38. Huang AS, Olson E and Moore DC. LCM: lightweight communications and marshalling. In: *2010 IEEE/RSJ international conference on intelligent robots and systems*, Taipei, 18–22 October 2010, pp.4057–4062. IEEE. DOI: 10.1109/IROS.2010.5649358.
 39. Kim KM. *Monocular visual odometry for fixed-wing small unmanned aircraft systems*. PhD Thesis, Air Force Institute of Technology, 2019.
 40. Cummins M and Newman P. {FAB-MAP}: probabilistic localization and mapping in the space of appearance. *Int J Robot Res* 2008; 27: 647–665.

Author biographies

Kyung Kim is a Cyberspace Operations Officer at Joint-Base San Antonio. He received his B.S. degree in Computer Engineering at the U.S. Air Force Academy in 2014 and his M.S. degree in Computer Science from the Air Force Institute of Technology in 2019.

Robert C Leishman is currently the Director of the Autonomy and Navigation Technology (ANT) Center at the Air Force Institute of Technology (AFIT) and a Research Assistant Professor in the Department of Electrical and Computer Engineering. He earned his M.S. and Ph.D. in Mechanical Engineering at Brigham Young University in 2009 and 2013 through support from the DoD SMART and USAF Palace ACQUIRE Scholarships. His research interests include robust alternative navigation, autonomous vehicles, computer vision, sensor fusion, and control.

Scott L Nykl is an Associate Professor of Computer Science at the Air Force Institute of Technology. His areas of interest are Real Time 3D Computer Graphics, computer vision, sensor fusion, parallel processing, Interactive Virtual Worlds, and computer networking. He received his B.S. degree in Software Engineering at UW-Platteville in 2006 and his M.S. and Ph.D. degrees in Computer Science from Ohio University in 2008 and 2013, respectively.