High-Altitude Vision-Based Navigation

For Future Mars Helicopter Missions

by Tanguy Gerniers

In fullfilment of the requirements for the Marshall Plan Scholarship Program

Home institution

Alpen-Adria Universität Klagenfurt

Institute for Smart Systems Technologies

Control of Networked Systems Group

Supervisors

Univ.Prof. Dr. Roland Brockers

Univ.Prof. Dr. Stephan Weiss

Host institution

NASA Jet Propulsion Laboratory,

California Institute of Technology

Robotic Aerial Mobility Group

Supervisors

Univ.Prof. Dr. Roland Brockers

Dr. Robert Hewitt

Acknowledgment

This research was carried out by

Tanguy Gerniers

at the

Jet Propulsion Laboratory, California Institute of Technology,

and was sponsored by the

JPL Visiting Student Research Program,

Alpen-Adria-Universität Klagenfurt,

Austrian Marshall Plan Scholarship,

and the

National Aeronautics and Space Administration.

(80NM0018D004)

Mentor

Univ.Prof. Dr. Roland Brockers

Task manager

Dr. Robert Hewitt

Robotic Aerial Mobility Group (347T)

Abstract

With its historical first flight in April 2021, the "Ingenuity" Mars Helicopter demonstrated the tremendous potential of using autonomous Unmanned Aerial Vehicles (UAVs) for planetary exploration. A key challenge that the NASA Jet Propulsion Laboratory (JPL) had to overcome to make this possible was finding a suitable navigation algorithm; as Global Navigation Satellite Systems (GNSS) are not available on Mars and most conventionally used navigation sensors are unpractical due to Size, Weight and Power (SWaP) constraints or Mars' unique geographic, geologic and atmospheric conditions. As a result, Ingenuity relies on the only other viable alternatives to date; visual, inertial, and range-based odometry (RVIO), to achieve sufficiently robust and accurate state estimation for safe aerial exploration. While Ingenuity's enduring success at low ranges, speeds, and altitudes exemplifies the viability of this approach, performance insights for the visual back-end of vision-based navigation algorithms outside these operational conditions are lacking in the existing literature. In particular, obtaining reliable and accurate state estimates during long and high-altitude traverses, such as those required to access important scientific targets in hazardous terrain and the planetary boundary layer, is expected to be challenging for a variety of reasons that are not fully understood. This work presents a survey of the performance, limitations and mitigation strategies for high-altitude vision-based navigation (VBN) in the current literature; and an evaluation of the performance and failure modes of the xVIO Range-Visual-Inertial Odometry Framework developed by NASA JPL, both in simulation and the real world.

Acknowledgements

This research was made possible thanks to the incredible individuals who shaped, supported, and inspired me before and during my stay at JPL.

First, I am immensely grateful to my supervisor Univ.Prof. Dr. Roland Brockers for the exceptional opportunity to join him in his research at the NASA Jet Propulsion Laboratory. His expertise in aerial robotics, open-mindedness, and good humor made for both an enjoyable and enriching experience.

Next, I extend my sincere thanks to JPL roboticists Dr. Robert Hewitt and Dr. Jeff Delaune for their invaluable support and wealth of knowledge, which were instrumental to this research. Special thanks also go to my fellow JPL interns Luca Di Pierno, Dario Pisanti, Ivan Alberico, and Moritz Reinders for their invaluable camaraderie and counsel.

Further, I am deeply appreciative of the support I have received from Univ.Prof. Dr. Stephan Weiss and Ph.D. Candidates Martin Scheiber, Christian Brommer, and Alessandro Fornasier of the University of Klagenfurt's Control of Networked Systems (CNS) research group throughout my master's studies. Their expertise in aerial robotics and excellent mentoring have laid a strong foundation for my academic and professional growth.

I would also like to acknowledge my sponsors, the Marshall Plan Scholarship and the University of Klagenfurt whose trust, generosity, and dedication to academic excellence made this research possible.

Finally, I would like to thank my family and my girlfriend, Ameline, for their unwavering love and support throughout my studies, for giving me the drive to dare mighty things, as well as the strength to see them through.

Contents

1 Introduction

1.1 The Mars Science Helicopter

The 2020 Mars Helicopter Technology Demonstrator (MHTD), better known as Ingenuity, was a huge milestone for aerial robotics that established autonomous, solar-powered Unmanned Aerial Vehicles (UAVs) as capable Mars explorers [1, 2, 3, 4]. Before Ingenuity had even completed testing however, NASA was already laying out plans for a larger and more capable successor; the Mars Science Helicopter (MSH) concept, capable of deploying up to 5kg of science instruments in previously inaccessible regions of Mars, such as the Valles Marineris canyon system and the planetary boundary layer (lowest 10 km of the atmosphere) [5, 6, 7, 8].

While Ingenuity's vision-based navigation algorithms have allowed it to operate well past its design specifications, reaching ranges, velocities, and altitudes of up to 704m, 10m/s, and 24m respectively [9], its reliance on a global planar ground assumption and the limited reach of its laser rangefinder have restricted its usage to relatively flat terrains and low altitudes [7]. Significantly improving on its predecessor, the MSH hexacopter aims to achieve traverses, cruise speeds and vertical profiles in excess of 10km, 30m/s and 2km respectively while maintaining accurate self-localization, thanks in part to the lower disk load of this design which will improve flight performance and controllability, the addition of advanced navigation algorithms such as Map-Based Localization (MBL) [10] and Range-Visual-Inertial Odometry (RVIO) [11], and the upgrade to a long-range laser altimeter [5, 6].

Figure 1.1: Mars Science Helicopter concept hexacopter configuration [5].

1.2 Motivations for high altitude navigation

High-altitude navigation capabilities are pivotal for Mars exploration, as they enable the investigation of a wide array of critical scientific questions. At only 120m AGL, a UAV can collect aerial imagery with comparable resolution to Mars rovers but at scales previously obtainable only from HiRISE satellite imagery [12, 13] while at 1km altitude, it can measure surface–atmosphere interactions in the Martian planetary boundary layer such as the exchange of heat, momentum, dust, and water, which are currently poorly understood and would revolutionize our understanding of the Martian atmosphere. Furthermore, access to certain remote science targets requires overflying complex terrain such as craters and canyon systems while maintaining a constant elevation (geometric altitude) to maximize endurance, which leads to significant variations in the altitude above ground level (AGL) [5, 6, 7].

However, flight at high altitudes AGL has direct implications for the accuracy of vision-based navigation, which tracks the translation and rotation between successive camera images to estimate changes in the UAV's pose. As the metric area covered by individual camera pixels, called Ground Sampling Distance (GSD), increases linearly with altitude AGL, so does the similarity between successive images in the absence of rotation, until the frame-to-frame camera translation becomes unobservable. Depending on the camera field-of-view (FOV), resolution, frequency, and velocity, this typically means that the position drift accumulated from the noisy accelerometer and gyroscope measurements becomes progressively harder to constrain by visual data the further away the camera is from the ground.

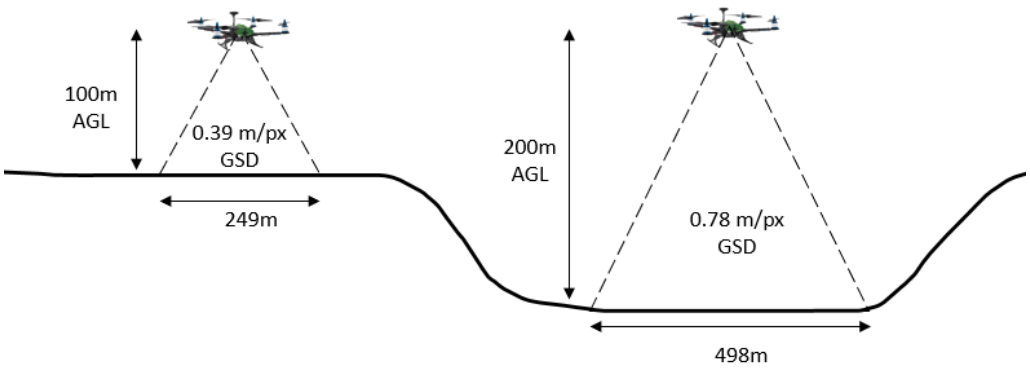


Figure 1.2: Image area coverage and GSD for a 640x480 OV7251 camera [14].

1.3 Research scope and outline

To fulfill the scientific and operational requirements mentioned above, a modular state estimation algorithm was motivated in [6] and developed to a limited extent in [7] at JPL for the Mars Science Helicopter. Leveraging laser altimeter measurements to enable low-drift self-localization in GNSS-denied non-planar environments, the xVIO range-visual-inertial odometry framework has shown promising results over relatively flat terrain at low altitudes [15, 16, 11, 17]. While additional research has been conducted to improve its initialization [18], heading estimation [16], and visual frontend [19, 20], little is known of its performance at high altitudes.

As part of the "Long Range Navigation and Autonomy" research task led by Dr. Roland Brockers at the NASA Jet Propulsion Laboratory, this work focuses on the evaluation and enhancement of xVIO's all-terrain access capabilities at high altitude over Martian terrain. In particular, we seek to evaluate xVIO's performance during autonomous system-in-the-loop (SITL) flights at 100m altitude using a monocular camera, an IMU, and a long-range laser altimeter.

Subsequently, this report is organized as follows: First, we review the current state-of-the-art in high altitude vision-based navigation including performance insights, limitations, and mitigation strategies. Next, we detail our research methodology, development, and limitations. Finally, we discuss our results and conclude with a brief recap and suggestions for future work.

For information regarding the notations and reference frames used in the following chapters, we refer our reader to chapter 2 of the original xVIO paper [15], having aligned our formulations with theirs wherever possible for consistency.

2 Literature review

In this chapter, we review the state-of-the-art (SOTA) in high-altitude vision-based navigation including suitable sensor suites, performance metrics and insights, key causes of drift and filter inconsistency, and corresponding mitigation strategies.

2.1 High-altitude vision-based navigation

In the context of vision-based navigation, defining what constitutes a "high altitude" is a complex task, as it varies significantly across the literature, between 30m and 2000m Above Ground Level (AGL) depending on the application. In the case of Range-VIO, high altitude could be considered as the upper quantile of the laser altimeter's maximum range, where we might expect a progressive degradation of the signal-to-noise ratio. In [17] for example, 100m AGL is considered high altitude given a 120m-range laser altimeter while in [21] it is 300m AGL given a 750m-range radar altimeter. For the particular case of Stereo-VIO, high altitude can describe the point at which it degrades to Mono-VIO, between 30-120m AGL according to [22, 23, 24]. For standard Mono-VIO however, the general consensus seems to be over 150m AGL for rotary UAVs [25, 26, 27], which corresponds to the minimum flight altitude for aircraft operation in the US [28] and the EU [29], while fixed-wing UAVs usually only receive that denomination past 2000m AGL [30, 26]. Taking these into consideration, we define "high altitude" as 100m AGL or higher for this work.

2.1.1 Sensors

As the name indicates, vision-based navigation requires visual data to permit accurate self-localization, in a process known as Visual Odometry (VO). This and other useful data can be obtained from sensors; active or passive devices capable of measuring physical phenomenons.

Active sensors, which emit signals to observe the environment, typically do not satisfy the Size, Weight and Power (SWaP) requirements for the MSH, which faces an extremely thin atmosphere with limited battery capacity. For instance, while Time-of-Flight (ToF) 3D LiDAR, ultrasound, radar and infrared sensors can measure distance and generate 3D Digital Elevation Maps (DEMs) using self-generated light, sound, radio and infrared waves respectively, they are too large, heavy, and power-hungry for Mars rotorcraft [31, 32, 21, 33]. Additionally, certain sensors like infrared depth cameras can be particularly vulnerable to solar radiation during daytime [16]. However, small and lightweight single-point LiDAR altimeters have proven to be reliable and effective against scale drift in the SOTA [11, 17, 34, 35, 36] and are therefore likely candidates for the MSH sensor suite.

Passive sensors, which measure naturally occurring phenomena, generally satisfy SWaP requirements for UAVs but can encounter problems in Mars' unique environment [37, 5]. For example, magnetometers and barometers are common UAV sensors on Earth but struggle to estimate absolute heading and elevation in Mars' weak magnetic field and thin atmosphere [5, 31, 16]. Similarly, Sun-sensors can reliably recover absolute orientation from the Sun position on Mars rovers [38, 39], but their reliance on accurate Martian time makes them vulnerable to clock resets at low temperatures without sufficient heating, as experienced by Ingenuity [40].

Inertial Measurement Units (IMU) are essential passive sensors for high-altitude navigation because their performance is not affected by altitude or terrain topography. Combining the linear acceleration measurements of an accelerometer with the angular rate measurements of a gyroscope, IMUs estimate the system's 3D position and orientation through (double-)integration, which makes errors accumulate, and the estimates thereby "drift" over time [31, 33]. While inclinometer pitch and roll data can help mitigate this drifting behaviour by initializing the IMU biases before take-off, cameras can both estimate inertial biases and constrain inertial drift throughout the flight [38, 36]. In particular, single-lens monocular cameras are best suited for this task as double-lens stereo cameras weigh more yet degrade to monocular setups at high altitudes [34, 41, 16], while shutterless event cameras are underrepresented in the literature and more complex to use [20].

The sensor suite considered for the MSH in [7] combines an IMU, a monocular camera and a laser altimeter; all with proven track-records on Mars [1] and at high altitudes [2.1.3] thanks to their low cost, size, weight, and power-usage [33].

2.1.2 Metrics

Measuring the accuracy of odometry systems in an objective, standardized manner, is important to understand their behavior and compare their performance. While denominations and formulations varies widely accross the literature, most equate to the popular ATE and RPE performance metrics.

The Absolute Trajectory Error (ATE), also called Absolute Pose Error (APE), is a global metric that quantifies in a single scalar the overall consistency of an odometry system across N filter propagations. Generally computed as the root mean square error (RMSE) between the estimated p_{est} and true positions p_{true}; usages of the mean, median and standard deviation also exist in the literature [21, 34, 37]. The ATE is often expressed as a percentage of the trajectory distance to facilitate comparison [42, 43, 44, 45, 17], with 0.1-2% of flight distance as ATE being considered the golden standard for VIO [46].

$$\text{ATE} = \sqrt{\frac{1}{N}\sum_{i=0}^{N-1} \|p_{est,i} - p_{true,i}\|^2} \tag{2.1}$$

Conversely, the Relative Pose Error (RPE) is a local metric that can be computed as the ATE for a specific time interval Δ rather than the whole flight, which is useful for isolating error peaks [43, 44, 42, 45, 47].

2.1.3 Benchmarks

Monocular high-altitude vision-based navigation is a recent and underrepresented topic in the literature. While comparing algorithms across different systems and environments is difficult, we find that most algorithms achieve less than 2.5 % of flight distance as ATE at high altitudes over urban or rural terrain, with noticeable improvements when using active ranging sensors [48].

Year	Paper	Odometry	ATE as %	Altitude (m)	Length (km)	Terrain
2012	[48]	VIO	2	500	0.5	urban
2012	[48]	VIO	10	500	0.5	alpine
2012	[48]	VIO	13	500	0.5	desert
2015	[21]	VO	4.67	600	16.5	rural
2015	[21]	Radar-VO	2.07	600	16.5	rural
2015	[34]	RVIO	0.78-0.39	150-300	4.5-7.8	rural
2022	[17]	VIO	1.4	100	0.5	urban
2022	[17]	RVIO	1.2	100	0.5	urban
2022	[17]	MSCKF-RVIO	1.02	100	0.5	urban
2022	[30]	VO	0.072-2	2530	78	rural
2023	[24]	VIO	2.5	100	0.8	urban

Table 2.1: Accuracy of SOTA mono-VO algorithms at high altitudes

2.2 xVIO Range-Visual-Inertial Odometry

Based on the popular Extended Kalman Filter (EKF) [49], xVIO [15] estimates the 9 DoF (position, velocity, and orientation) of the nonlinear UAV system, the IMU biases, and the filter's confidence in these estimates, generally referred to as the state uncertainty or covariance. Integrating inertial measurements over time, xVIO predicts new estimates of the state at high frequency, then updates and corrects these estimates at a lower frequency using visual data from a monocular navigation camera and range data from a laser altimeter, in a process known as range-visual-inertial odometry (RVIO).

xVIO performs visual updates using the FAST feature detector [50], the pyramidal implementation of the Kanade-Lucas-Tomasi feature tracker (KLT) [51, 52], and 2 types of visual update paradigms; Simultaneous Localization And Mapping (SLAM) and Multi-State Constraint Kalman Filtering (MSCKF) [53]. However, as MCSKF cannot triangulate features with sub-pixel parallax at higher altitudes [19], we focus on the SLAM paradigm in the remainder of this work.

In its range update, xVIO leverages its existing SLAM features to perform a Delaunay triangulation [54] of the terrain, and makes a local planarity assumption for the triangle currently intersected by the laser altimeter to approximate the inverse-depth [55] of the 3 features that make up its corners, using the range measurement. This allows the filter to recover 3D information from the 2D image and thus constrain scale drift, a key problem in VIO [56].

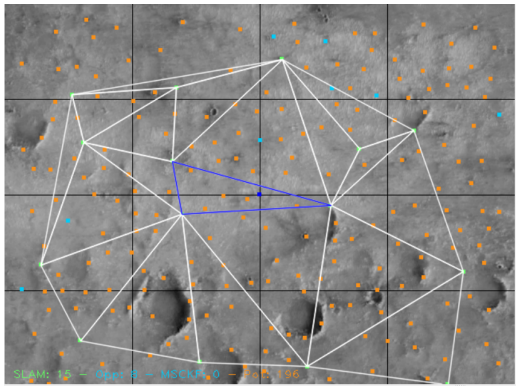


Figure 2.1: Visualization of a xVIO range-visual update in *Gazebo*. In green are 15 visual state SLAM features, in blue the laser altimeter ray and the intersected Delaunay triangle, and in cyan/orange currently unused potential SLAM features.

2.3 Limitations & mitigation strategies

2.3.1 In extended Kalman filtering

Vision-based navigation systems are dynamic and highly nonlinear, due to the intricate relationships between the system, the sensor measurements, the environment, and time. Consequently, state estimation for such systems is sub-optimal, as we can only approximate future states and their uncertainty. In EKFs, estimation is done by analytically linearizing the dynamic and measurement models of these systems around their Gaussian mean, at each point in time, using first-order Taylor series expansions. While easier to model and least memory intensive, they assume local linearity around the current state estimate and tend to be heuristic in nature, focusing only on a few critical and well-defined relationships in the interest of efficiency. This leads to linearization errors, making EKF state estimates only as accurate as their models are comprehensive and locally linear. However, these errors can be mitigated by using Iterative EKFs (IEKF), which iteratively refine state estimates around which the linearization is performed, and Unscented KFs (UKF), which bypass the local linearity assumption using statistical linearization [15, 49, 57, 58].

Furthermore, EKFs are only consistent and therefore reliable, while their model errors are fully characterizable by a Gaussian mean and an uncertainty, represented by the estimated covariance matrix. In reality however, many VIO processes such as feature tracking and nonlinear dynamics produce non-Gaussian error distributions, leading EKFs to frequently underestimate the true covariance matrix and risk becoming inconsistent [59]. Once inconsistent, EKF estimates typically diverge rapidly, forcing the UAV to perform an emergency landing in the absence of mid-air initialization capabilities. This risk can be mitigated through covariance inflation techniques like the addition of "stabilizing noise" to the process or measurement noises to better reflect the actual uncertainty in the system, or statistical filtering methods like UKFs to circumvent risks associated with mismodelling, local nonlinearity and non-Gaussian noise [60].

For filter-based VIO methods such as xVIO, the initialization of the filter is the first and most critical step, as this is where the initial estimates for the analytical linearization process are defined or estimated. While errors in the initial x-y position will only induce a constant position error, errors in the initial yaw will result in uniform position drift over time, and errors in the initial pitch, roll, gravity vector, or uncertainties may lead to rapid divergence. Accurately estimating these parameters is a challenging task due to the number of variables and poor conditioning of the system [37], that is typically done using calibrations and absolute localization data before takeoff [15, 16, 38], or through self-initialization techniques such as closed-form solutions [61, 62], spline-fitting [37, 63] and bundle-adjustment [18].

2.3.2 In inertial odometry

As mentionned in [2.1.1], inertial odometry is prone to drifting over time due to the (double)-integration of inertial noise, which stems from various sources. While all IMUs induce small quantization errors due to their limited floating-point precision, low-cost COTS IMUs such as those used in current and planned Mars rotorcraft typically have significant biases and noises due to manufacturing imperfections and electromagnetic, cross-axis and temperature sensitivity. During flight however, turbulence from backwash or strong winds and mechanical vibrations from the pro-

peller(s) can become the main sources of inertial noise without proper dampening or covariance inflation [64, 34]. These noises, and an other unmodelled perturbators of the system dynamics, are ideally accounted for by the zero-mean process noise of an EKF, which should be carefully balanced; too much process noise and outliers will bias the estimates, too little and important inliers will be ignored [58].

With sufficient excitation and a well-defined process noise, an EKF will drift minimally. If certain axes are not or insufficiently excited however, they become unobservable for the filter, which is incapable of differentiating noise from near-constant measurements. While hovering or during traverses at constant velocity for instance, IO consequently drifts in global position, velocity, and orientation, but remains observable in pitch and roll thanks to their relative orientation to the gravity vector [65, 37, 66]. Nevertheless, unobservability in the x-y position axes, velocity and attitude can be mitigated using visual data, and in the z position axis using range data [11].

2.3.3 In visual odometry

Cheap and versatile, visual odometry is an essential part of most robotic exploration systems whose performance depends on 4 key components; quantization error, tracking error, measurement noise, and observability; which respectively define the precision, accuracy, weight and informativeness of visual updates.

2.3.3.1 Quantization error

Quantization refers to the discrete pixel-precision at which features can be tracked, and by extension the metric-precision of the visual update, represented by the Ground Sampling Distance (GSD) in meters per pixel. From the formulations of the GSD in [67, 68, 69], we intuitively observe a linear relationship between the altitude AGL and the feature tracking precision, whereby for a given height AGL and level of process noise, increasing the camera resolution should decrease the position error and uncertainty by improving the precision with which the visual update corrects the prediction. Conversely, increasing height AGL for a given camera resolution should increase the quantization error and by extension also the position

error and uncertainty, although this couldn't be proven by [24] within the scope of their high-altitude mono-VIO experiments.

$$\text{GSD} = \frac{\text{altitude AGL} \cdot \text{sensor size}}{\text{focal length} \cdot \text{image size}} \tag{2.2}$$

To mitigate the loss of precision resulting from high altitudes and poor camera resolutions, image enhancement techniques such as Super-Resolution (SR) are popular, since larger camera sensors typically come at the cost of more pixel noise. While Single Image SR (SISR) techniques, which upscale individual low-resolution images have been successfully applied to VIO [70], their generative nature can be seen as an unnecessary additional source of uncertainty and tracking error. Multi-Image SR (MISR) solves this problem by combining multiple low-res images into a single high-res output that ensures reliable pixel registration [71, 72].

2.3.3.2 Tracking error

The tracking error represents the feature tracker's ability to accurately track features across consecutive images up to a given precision. For indirect, descriptor-based feature trackers such as SIFT, SURF, and ORB, which are relatively robust to changes in scale, rotation, and lighting; this mostly depends on the quality of the feature descripors and matching methods used. Direct, intensity-based feature trackers on the other hand, such as the Kanade-Lucas-Tomasi tracker (KLT), are faster and more robust to scene homogeneity, but sensitive to a wider variety of parameters. Here, we focus on the limitations of the pyramidal KLT used in xVIO.

To begin with, the KLT algorithm assumes feature motions to be locally linear and "spatially coherent", as it finds feature correspondences in consecutive images by perform gradient descent on the pixel intensities around predicted feature positions [73]. This makes KLT vulnerable to feature displacements that exceed the frame and tracking rate, such as extreme accelerations and pure rotations, which place predictions outside the image and thus make correspondence impossible, forcing the filter to linearize between fewer matches [37]. However, this risk can be mitigated by increasing the image rate or using gyro-based derotation to overcome large attitude

motion [74]. Another limitation of this assumption is that because predicted feature positions are always computed from priors, inherent errors in the prediction due to imperfect camera calibrations and distortion models accumulate and induce position drift over time, albeit less than IMUs. Nevertheless, these prediction errors can be minimized through rigorous calibration and the use of appropriate distortion models, such as the FOV model [75] for wide-angle monocular cameras, while the inevitable drift can be attenuated by using inertial data in the prediction step to improve the accuracy of the priors and reduce computation time [66].

Next, since KLT prioritizes which features to track based on their quality, more features does not necessarily guarantee better results [66]. While most natural landscapes exhibit fractal properties, such that the quantity and quality of features are invariant with scale [76], researchers have observed a relationship between scale and image blur which decreases the size and number of feature points at higher altitudes [70, 77]. To combat this effect, the pyramidal implementation of the KLT can track at lower resolutions to facilitate convergence at every scale [51].

Then, because KLT uses the gradients of image intensities to estimate motion, it's performance depends on the ambient illuminance and the terrain albedo. For instance, while flying in overly or insufficiently bright environments, in poor visibility conditions, over homogeneous terrain, or over terrain with overly low or high reflectance, pixel intensities tend to me homogeneous across the image, making features more ambiguous and convergence more difficult, as gradient descent faces smoother pixel intensity slopes and more local minima [46, 48, 24]. To mitigate this, [74] used an autoexposure algorithm on the Mars Helicopter to optimize image intensity for feature tracking across different terrains and weather conditions.

Lastly, while the ISO analog gain and exposure time settings of monocular cameras can be minimized to limit image noise and motion blur, and a maximum exposure level can be enforced to satisfy the constant brightness assumption of the KLT's gradient-descent algorithm [74], limitations remain for the velocity, turn rate, and altitude AGL, to keep motion blur below a tolerance level of a couple of pixels and ensure sufficient image overlap for feature correspondence [78, 65, 79, 21].

2.3.3.3 Measurement noise

The visual measurement noise is the zero-mean pixel-variance that encompasses both the tracking and quantization errors, which combined typically range between 0.33 [16] and 1 pixel σ [80] for Mono-VIO. Set too high, the filter won't trust the visual measurements and will face unconstrained inertial drift. Set too low, and the filter might become overconfident, allowing outlier measurements to corrupt the state and lead it to diverge, or end up in "lockout" if the state estimates are sufficiently perturbed for good visual measurements to become indefinitely rejected [36].

2.3.3.4 Observability

Firstly, the true scale of the scene is unobservable in Mono-VO without motion constrains or prior knowledge of the environment as 2D images provide no depth information, such that fast motions at high altitude cannot be distinguished from slow motions at low altitude. Consequently, VO depth and pose estimates are only accurate up to an unknown scale factor, which drifts over time [30, 33, 34, 81, 82]. Fast vertical motions at low altitudes and slow horizontal motions and high-altitude are especially likely to induce scale drift, as even the estimated scale factor becomes unobservable in excessive [21] and sub-pixel [22] feature translations. This is particularly dangerous for EKFs, as nonlinear visual measurement models are sensitive to noise and may converge to local minima and therefore inconsistent solutions during takeoff, landing, or at high altitudes [53]. To this end, a real-time scale drift estimation algorithm was recently developed by [83] to help UAV autonomy algorithms monitor and incorporate this risk in their decision-making.

Another problem with scale unobservability is that the depth of features is unknown at the time of their initialization. In the absence of MSCKF or to improve the stability of MSCKF triangulation when baselines are insufficient, xVIO initializes the depth of new SLAM features using a constant inverse-depth $\hat{\rho}_0 = \frac{1}{2d_{min}}$ and uncertainty $\hat{\rho}_0 = \frac{1}{4d_{min}}$ that are defined before filter initialization based on a parametrizable d_{min} minimal depth that defines a 95 % acceptance region for depths between d_{min} and ∞ [15]. The inverse-depth parametrization is used here, since visual features could theoretically have infinite depth [55]. As poor initial

feature depth estimates can lead to unconstrained scale drift however, alternative approaches were explored in [17] to initialize the depth and depth covariance of SLAM features using a laser altimeter range measurement, which account for the pixel-wise distance to the LRF intersection point and the range measurement noise.

VO techniques exist to mitigate scale drift, such as Simultaneous Localization and Mapping (SLAM) or Structure from Motion (SfM) algorithms which estimate scale over time through triangulation, or deep learning scale estimation networks [84, 56, 30], but these are memory-intensive and challenging to tune for performance [56, 81]. Flat world assumptions are a common alternative that work well for flights with low topography-altitude ratios, but are incompatible with Mars's rough landscape and the MSH's comparatively low flight altitudes [77, 66, 33]. Consequently, VIO is often used to recover the true scale of the scene [37], but fails to do so when hovering or flying at uniform velocity [16, 85], and degrades in situations with larger inertial-visual excitation ratios as visual measurements become less informative due to insufficient parallax and ergo cannot constrain inertial drift, especially if the measurements are noisy and the filter is not consistent enough to extract all the information it needs from the image [17, 37, 41].

At high altitude, it has been highlighted by [66, 81, 21] that translations becomes paradoxically more difficult to estimate because while feature tracks get longer and features ergo have more time to be triangulated, parallax decreases and translations therefore become more difficult to observe. In their experiments at different speeds and high altitudes however, [48] did not find these effects to be significant. When dealing with VIO specifically, [37, 63] also find that visual odometry's requirement for sufficient parallax at high altitudes result in excessive amounts of mostly uninformative IMU measurements and thus significant drift due to long periods of unexcited inertial integration and bias unobservability.

Secondly, Mono-VIO cannot observe global heading about the local gravity vector in the absence of rotation [15, 86, 16, 87]. This is a significant problem for navigation because small heading errors can result in significant position drift over long distances, often making them the primary source of position error, as was the case for Ingenuity's first flights [88]. Visual compass algorithms can help mitigate this

problem by indirectly recovering the system's yaw from the position of the sun by tracking the UAV's [89] or the scene's [90, 91, 92] shadows, or by estimating changes in heading relative to the geomtry of the terrain [93, 65, 94].

Finally, [66] observed an "attitude coupling problem" in mono-VIO while hovering, whereby drift in pitch resulting from gyroscope noise is interpreted as a translation by the filter due to the use of attitude in the visual measurement equation. In particular, they found the resulting error to have an oscillatory behavior that increases linearly with altitude and results in divergence above 700 meters.

2.3.4 In range-visual odometry

To overcome the scale ambiguity problem in VIO, the fusion of range measurements from a laser altimeter was proposed in [11] and further explored in [17], demonstrating reductions in scale drift up to a factor of 5 and in position error up to a factor of 9 at low altitudes, with a certain level of tolerance to terrain roughness. The range update's ability to recover the scale of the scene depends on several key factors:

First, the Signal-to-Noise-Ratio (SNR) of the laser altimeter measurement defines the accuracy of the metric depth information VIO needs to observe its scale. According to [95, 96], a laser rangefinder's SNR linearly depends on a variety of factors such as the distance travelled by the signal, the signal beam divergence angle, and the terrain's reflectance and slope, all of which should be minimized, save for the reflectance, to maximize the SNR and thus the measurement's accuracy. [97] seperates these factors into range-independent noise, such as potential sensor biases due to manufacturing defects, and range-dependent noise, such as errors from averaging over larger spot areas due to beam divergence over long distances.

Second, the LRF orientation and x-y translation w.r.t the camera define which image pixel, and therefore which Delaunay triangles the LRF will intersect. Because xVIO assumes no LRF-camera translation however, any offset in z directly translate to error in the range measurement, then in the state estimates [11]. Offsets in x or y can on the other hand be neglected at high altitudes, as their impact become sub-pixel due to the GSD. To mitigate this and changes to the calibration during flight, [17] perform online extrinsic calibrations for the LRF and camera.

Next, the range update makes the assumption that the terrain within Delaunay triangles is locally planar, since most visual features mark depth discontinuities and the area between them is typically smooth [11]. Extending this work, [17] include the features of co-planar neighboring triangles to the range update to further disambiguate the scale at high altitudes, and use the distance measurement to initialize the inverse-depth and covariance of features that are co-planar with the base triangle, based on their distance to the LRF intersection point and the range measurement noise. While [11, 17] demonstrated in their experiments that the 19 triangular facets produced by 15 SLAM features could efficiently capture the structure of 3D scenes up to 110m, common terrain features on Mars such as craters and canyons, and the increased facet area at high altitudes, are likely to stress these assumptions. Observing these pitfalls at very high altitudes, Ivan Alberico (2023) adapted xVIO to prioritize features passing through the LRF intersect pixel above other potential SLAM features with higher FAST scores, such that through the cross-correlations of the state matrix, the depth of all features become observable from the known metric depth of this new feature.

Then, the measurement noise of the range update is the zero-mean meter-variance that encompasses both the altimeter's SNR, calibration errors, errors in the feature depth estimates, and violations of the local planarity assumption. Used in the Mahalanobis distance test to reject outliers, this filter is difficult to tune for rough terrain and high altitudes because at the levels required to avoid outliers from the aforementioned effects, the update's weight becomes negligible. To circumvent this problem, [36] individually tune the covariance of their different axes to optimize the informativeness of their measurements, while [98] use an improved adaptive EKF (AEKF) to adapt the measurement covariance in real time.

Finally, for the scale to be observable in the absence of excitation, RVIO also requires translation motions to be observable such that feature depths are known and correlated. Similar constraints are found in the work of [21] on Radar-VO. However, this parallax requirement is not satisfied while hovering or at high altitudes [15].

3 Methodology

The following chapter details how this research was conducted starting with the overall design, before diving deeper into how the data was collected, analyzed and visualized, and concludes with a reflection on the limitations of these methods.

3.1 Research Design

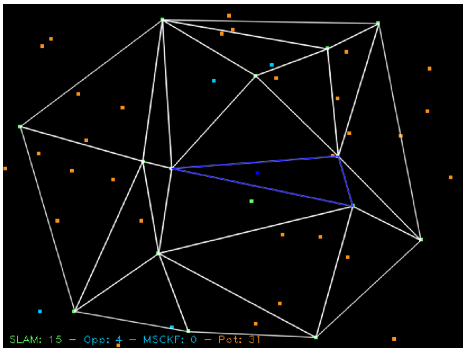
Considering that this research aims to evaluate the performance and failure modes of xVIO, and the viability of potential mitigation strategies, this work adopts an experimental research design [99] to evaluate the impact of various parameters and mitigation strategies identified in the literature review, on the xVIO state error and uncertainty. The hypotheses we define for each independent variable are tested in Monte Carlo simulations to ensure the statistical significance of the results; both in a realistic robotics simulator to obtain results as close as possible to reality using a graphics engine, physics engine, and flight controller, and in a custom heuristic simulator free of tracking errors from the feature tracker or graphics engine, and integration errors from the physics engine or flight controller. Finally, we conduct high-altitude outdoor flights to evaluate the performance of xVIO in the field.

3.2 Data Collection

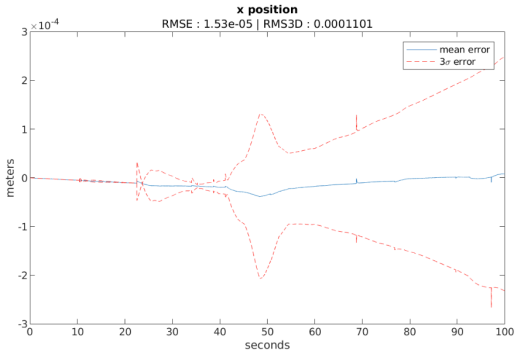
3.2.1 Heuristic Monte Carlo simulations

To begin with, a heuristic simulation framework using simplified control and sensor models was required to test our hypotheses in a controlled environment, free of the uncertainty that surrounds graphics and physics engines, and tracking and control algorithms. While the purely analytical MATLAB-based range-visual-inertial simulation environment developed by JPL and used by [82] fulfilled these requirements, its slow runtime, high memory usage and poor integration with the data analysis

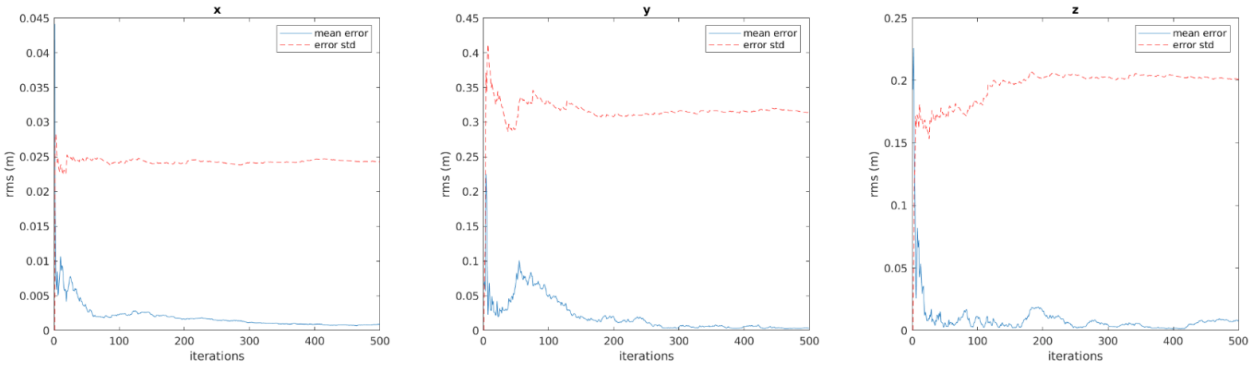
software made it ill-suited to run our numerous Monte Carlos. Consequently, we developed a new heuristic simulation framework written entirely in C++. Capable of generating RVIO sensor measurements and ground truth data more than 32 times faster than its predecessor when multithreading, xSIM-Cpp also features an improved trajectory interpolation algorithm, 3D-terrain compatibility for both visual and range measurements, automatic LRF calibration, and lossless recording to the rosbag format. The simulation framework was validated with xVIO using zero-noise runs, resulting in only floating-point errors across kilometer-long trajectories as seen in figure 3.1b, and the sample size for each trajectory was determined by the convergence point of the position error mean and variance, which roughly happens around 300 iterations in figure 3.1c.



(a) Visualization of an xVIO update using SIM-Cpp sensor data



(b) x position error across a zero-noise 1km traverse at 100m AGL

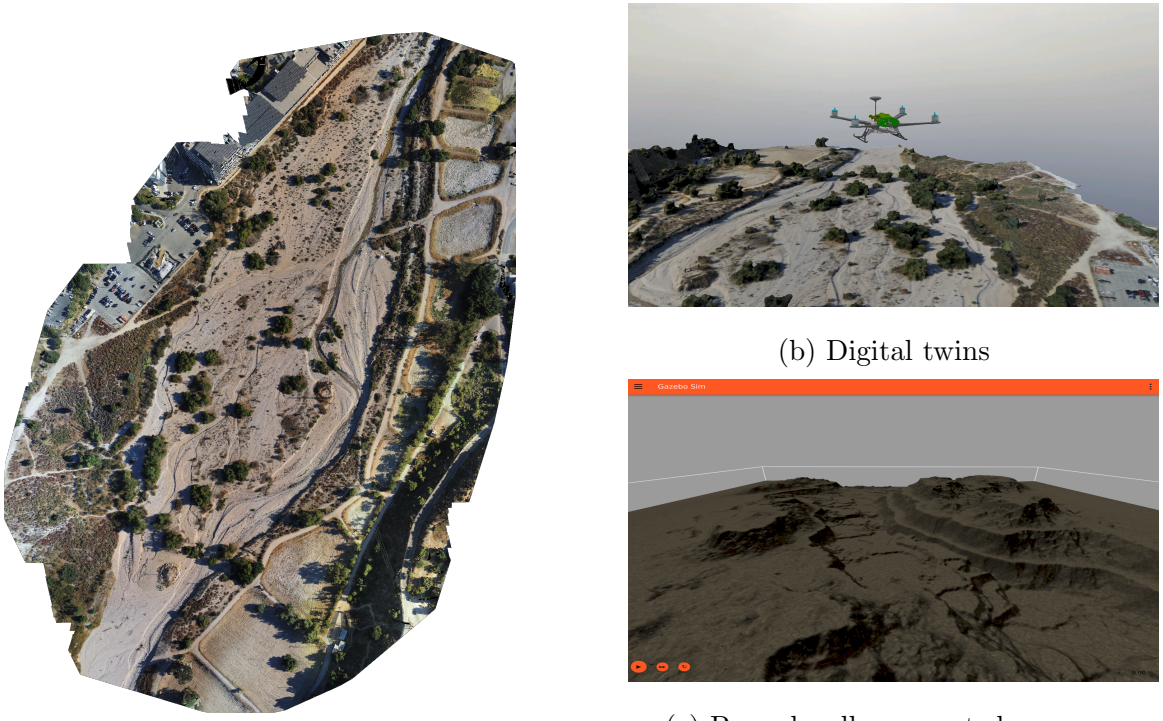


(c) Position error mean and variance stabilization with sample size

Figure 3.1: xSIM-Cpp methodological considerations

3.2.2 Realistic Monte Carlo simulations

Next, a realistic simulation framework was required to evaluate the impact of tracking and control algorithms as well as real 3D terrain models on our hypotheses. We first considered AirSim [100], ETH flightmare [101] and VINS Eval [102]; popular open-source 3D robotics simulators built on the Unreal and Unity game engines and capable of rendering highly realistic environments, but ultimately decided against due to their unpredictable rendering artifacts and high memory usage which made simulations unstable. JPL's proprietary Darts/Dshell multibody simulation framework was also considered, having being previously used to study Ingenuity's flight dynamics [2], but its steep learning curve, outdated graphics engine and lack of ROS and PX4 integration made it unsuitable. Consequently, we opted for the latest version of Open Robotics' popular and lightweight open-source 3D robotics simulator "Gazebo" [103, 104]. Offering native ROS and PX4 integration, numerous sensors and physics engines, long-term support, and photorealistic environments, this simulator was recently used at JPL by [105] to perform RVIO System-In-The-Loop (SITL) simulations with a quadcopter in Mars lava tube environments.



(a) Aerial map

(b) Digital twins

(c) Procedurally generated canyon

Figure 3.2: Gazebo visual models

To replicate our real outdoor flights in Gazebo using the recorded flight plans, digital twins of the UAV and flight environments as seen in figure 3.2b were created to scale using aerial 3D terrain reconstructions as illustrated in figure 3.2a, and CAD models of the UAV frame and sensors. Similarly, digital twins of actual Martian terrain were created using HiRise Digital Elevation Models (DEMs) and aerial imagery [12]. For hypotheses requiring rougher terrain however, we procedurally generated Mars-like terrains with craters and canyons using Blender, as illustrated in figure 3.2c. Because we used the same PX4 flight controller for this simulation framework as for our outdoor flights, the GZPX4-MC autonomy algorithm also had to be developed to synchronize PX4 with the Gazebo simulator and the xVIO state estimator via ROS, such that Monte Carlos could be run autonomously.

For results from the Gazebo simulation to be comparable with those from xSIM-Cpp, we replaced Gazebo's IMU noise model with our simplified model detailed in *Appendix A*, and replaced Gazebo's radial image distortion model with an implementation of the FOV distortion model [75], which can be found in *Appendix B* and is visualized in figure 3.3 using the OGRE NEXT graphics engine.

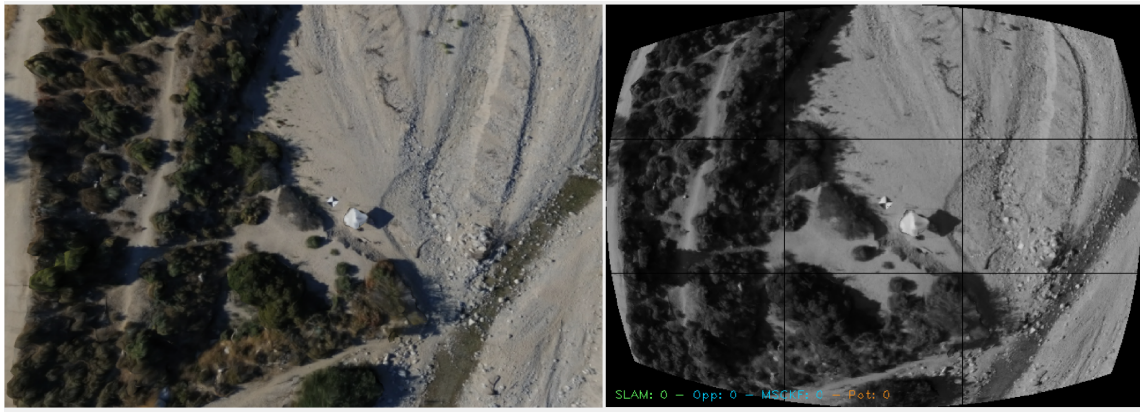


Figure 3.3: Distortion of a raw Gazebo image (left) using the FOV model (right)

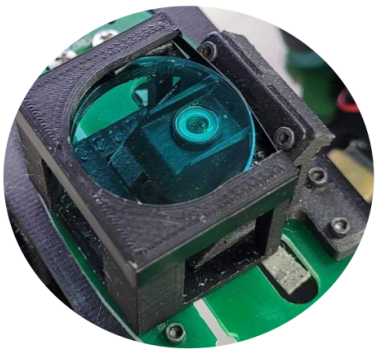
To ensure its validity, this simulation framework was sanity-checked and had its trajectory sample sizes defined in the same manner as xSIM-Cpp. However, due to the added uncertainty of the feature tracker, PX4 controller, ODE physics engine and 3D terrain, a small position error is inevitable even in the absence of measurement noise, and the required sample size was subsequently much larger.

3.2.3 Outdoor closed-loop flights

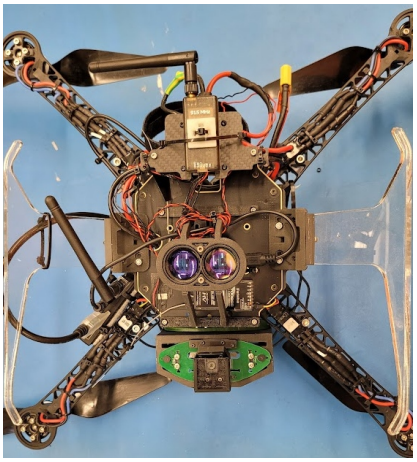
Finally, outdoor closed-loop flights were required to evaluate xVIO's high-altitude performance in practice. To this end, a ModalAI Sentinel quadcopter equipped with a VOXL2 autopilot, a TDK ICM-42688-P 6DoF IMU, and a nadir-pointed Omnivision OV7251 wide-angle global-shutter 640x480 monocular navigation camera, was augmented with a nadir-pointed Lightware SF30D 200m-range laser altimeter to match the sensor suite considered in [7] for the MSH. Additionally, the Sentinel's integrated magnetometer and GNSS antenna were preserved to provide ground truth outdoors, and the UAV frame was augmented with 6 reflective sensors to enable initial SITL tests indoors using a VICON motion-capture system which we set up for this purpose. For SITL to be possible however, an additional Pixhawk4 Mini autopilot had to be installed, as the EKF2 estimator of VOXL2's integrated PX4 version could not fuse external pose estimates from xVIO. Also, a UART- and USB-compatible ROS driver had to be developped for the SF30D laser altimeter due to its novelty, and was later upgraded with self-reinitialization capabilities to counteract occasional signal cut-offs due to vibrations during flight.

After passing initial manual SITL tests with xVIO inside the JPL micro aerial vehicule laboratory, autonomous waypoint-following SITL tests were performed using PX4 at low altitude in the JPL Mars Drone Yard seen in figure 3.4a, then at higher altitudes in the Arroyo Secco creek shown in figure 3.2a.

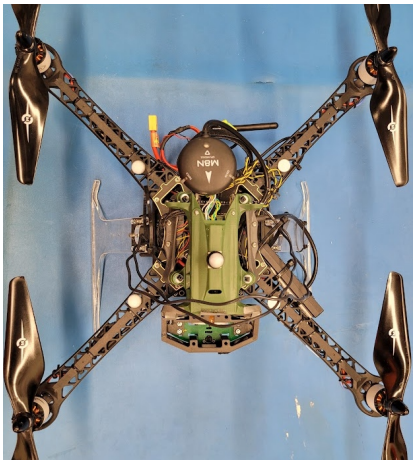
(a) JPL drone yard



(b) Camera with bandpass filter



(c) Sentinel bottom view



(d) Sentinel top view

Figure 3.4: Experimental setup

During testing, we discovered serious over-exposure problems with the camera when flying at noon with a clear sky over the mostly sandy and rocky terrain of the Arroyo Secco creek, both with autoexposure enabled and manually setting the exposure time to its minimum value of 4ms. Consequently, we used an IR-cutoff filter or a Bandpass colored-glass filter depending on the lighting conditions to mitigate this problem, and developed custom removable mounts as seen in figure 3.4b for these that minimally obstruct the camera.

(a) Default (b) With IR-cutoff filter (c) With bandpass filter

Figure 3.5: Image saturation with autoexposure in the JPL Mars Yard at noon

Before each flight, the IMU and camera intrinsics and extrinsics were calibrated using the Kalibr visual-inertial calibration toolbox [106] and a calibration board, and the LRF intersect pixel point was defined from an image of a 10m-distant flat surface. The LRF intrinsics (noises and biases) were calibrated once for all flights, and are explained in more detail in the next chapter.

3.3 Data Analysis

The sensor data we collected was evaluated in real-time by xVIO using a custom C++ ROS wrapper, whose state estimate outputs where then compared to the ground truth (from the simulator or GPS) using various MATLAB scripts which also analyzed other factors such as update residuals. While most VIO literature expresses state estimation accurary solely in terms of ATE or RPE, our evaluations consider 4 different variations of the underlying Root Mean Square metric to measure the representativity of our samples (RMSE), the effective pose uncertainty (RMS3D), the estimated pose uncertainty (RMScov) and the effective measurement uncertainty (RMSgamma).

Let us assume an inertial state without biases x_I, its covariance δx_I, and their diagonals x_I^{diag} and δx_I^{diag} :

$$x_I = \begin{bmatrix} {p_w^i}^T & {v_w^i}^T & {\theta_w^i}^T \end{bmatrix}^T \tag{3.1}$$

$$\delta x_I = \begin{bmatrix} {\delta p_w^i}^T & {\delta v_w^i}^T & {\delta \theta_w^i}^T & {\delta b_g}^T & {\delta b_a}^T \end{bmatrix}^T \tag{3.2}$$

$$x_I^{diag} = \begin{bmatrix} p_{x_w}^i & p_{y_w}^i & p_{z_w}^i & v_{x_w}^i & v_{y_w}^i & v_{z_w}^i & \theta_{x_w}^i & \theta_{y_w}^i & \theta_{z_w}^i \end{bmatrix} \in \mathbb{R}_+^{1\times 3} \tag{3.3}$$

$$\delta x_I^{diag} = \begin{bmatrix} \delta p_{x_w}^i & \delta p_{y_w}^i & \delta p_{z_w}^i & \delta v_{x_w}^i & \delta v_{y_w}^i & \delta v_{z_w}^i & \delta \theta_{x_w}^i & \delta \theta_{y_w}^i & \delta \theta_{z_w}^i \end{bmatrix} \in \mathbb{R}_+^{1\times 3} \tag{3.4}$$

The error between the estimated and actual inertial state can be written as :

$$\Delta x_I^{diag} = x_{I,est,i}^{diag} - x_{I,true,i}^{diag} \tag{3.5}$$

For N inertial propagations across a trajectory, and M iterations of this trajectory in a Monte Carlo sequence, the performance metrics can therefore be written as :

$$\mathrm{RMSE}_{(N,M)} = \sqrt{\frac{1}{N}\sum_{i=0}^{N-1}\left\|\frac{1}{M}\sum_{i=0}^{M-1}\Delta x_{I,i}^{diag}\right\|^2} \tag{3.6}$$

for the Root Mean Square Error, which measures the population sample bias and

therefore converges for all states of a consistent filter towards 0 as M increases.

$$\mathrm{RMS3D}_{(N,M)} = 3 \cdot \sqrt{\frac{1}{N}\sum_{i=0}^{N-1}\left\|\frac{1}{M}\sum_{i=0}^{M-1}\left(\Delta x_{I,i}^{diag} - \frac{1}{M}\sum_{i=0}^{M-1}\Delta x_{I,i}^{diag}\right)\right\|^2} \tag{3.7}$$

for the Root Mean Square 3-Sigma variance, which measures the actual pose uncertainty across the trajectory, such that each point of the RMS3D curve encompasses 99.7% of the errors encountered at that timestamp across all M Monte Carlos.

$$\mathrm{RMScov}_{(N,M)} = \sqrt{\frac{1}{N}\sum_{i=0}^{N-1}\left\|\frac{1}{M}\sum_{i=0}^{M-1}\delta x_{I,i}^{diag}\right\|^2} \tag{3.8}$$

for the Root Mean Square Covariance, which measures the estimated pose uncertainty of the Kalman filter across the trajectory, such that each point of the RMScov curve represents average metric confidence of the filter in a given state at that timestamp, across M Monte Carlos.

Following the same logic, we also evaluate the estimated measurement uncertainty (RMSgamma) for the visual and range updates, the average number of visual updates that pass the outlier rejection test (RMS_nb_inliers), the average pixel residual in the x (RMS_x_mean) and y (RMS_y_mean), and the average altitude of features in the world frame for visual (RMS_slam_z_mean) and range (RMS_lrf_z_mean) updates respectively.

Using these performance metrics, hypotheses are evaluated by comparing changes in performance with our perturbations of the independent variables, using the experiment configuration detailed in *Appendix C* as the reference point. Taking figure 3.1b as an example; the blue curve represents the RMSE, the red curves represent the distribution of the RMS3D around the RMSE, and their metric value (y-axis) across the trajectory duration (x-axis) are noted in the plot subtitle. For such a plot and large Monte Carlos of realistic data, the distance of the blue line from the x-axis would therefore represent the sample bias, the distance of the red lines from the blue line would represent the metric state error, and the smoothness of both serves as an indicator of the consistency of the filter.

3.4 Limitations and considerations

We conclude this chapter with a discussion of the limitations of our methodology, and certain considerations to account for when reviewing our results. First, the results originating from our heuristic simulation framework can be considered as inherently optimistic, as strong assumptions are made in how the data is generated to isolate the effects of the independent variables we vary from those of the feature tracker and flight controller. Next, due to the long computation times required to run our complex simulations, especially those requiring memory-intensive graphics and physics engines, the number and range of independent variables and mitigation strategies we evaluated, as well as their validation using Gazebo, was limited by the time we had available to conduct this research and the computational capabilities at our disposal (20-core Intel i5 13600KF CPU, Nvidia RTX 4060 GPU, 32GB RAM). For this reason, our results mainly focus on the transient phase of the nonlinear EKF system, as significant computational power is required to evaluate the long-term steady state, particularly when perturbing the system with additional sources of uncertainty such as high altitudes and 3D terrain. Additionally, because we used Earth's gravity for our simulations to match the parameters of our outdoor flights, the flight dynamics generated by the PX4 controller in our Gazebo simulations and outdoor experiments likely differ from those on Mars. Furthermore, the number, complexity, duration, and altitude of our outdoor experiments were limited by the capabilities of the UAV, the availability of drone pilots and observers, and the budget of the research task. Also, because the ground truth for our outdoor flights was provided by GPS, our results should only be considered accurate up to the precision of our GPS, which we estimate to be around 2m in all axes. Finally, it should be noted that the writing of this report was facilitated by the proofreading and paraphrasing functionalities of the Grammarly 2.5 AI writing assistant and ChatGPT 3.5 AI chatbot, of which the latter was also greatly helpful in bug- and problem-solving at various stages of the software development.

4 Summary of results

In this chapter, we summarize the outcomes of our analyses and experiments. We refer the reader to the master thesis (Tanguy Gerniers, 2024) that follows this report for the visualizations and in-depth discussions of our results.

4.1 Analyses in simulation

We begin with an analysis of the filter convergence properties, where we find that xVIO goes through a 3-step transient phase following initialization before settling into steady phase once the first tracked SLAM features leave the image, with uniform position error growth and cubic uncertainty growth driven primarily by scale drift, since visual updates are only able to constrain position drift up to a drifting scale factor. In the first step of the transient phase, unconstrained inertial drift leads the pose error and uncertainty to grow exponentially as the inertial biases are yet unknown. Once xVIO starts tracking SLAM features, the filter begins to approximate these inertial biases and thus constrain the inertial drift, which stabilizes the pose error and covariance. Finally, the pose error and uncertainty decrease as the filter converges on its best approximation of the inertial biases and recursively corrects its state and covariance.

Here, we find that while the system always converges to a linear position drift in the steady phase, both the filter convergence time in the transient phase and the position drift rate in the steady phase increase with altitude, resulting in a radical relationship between altitude and position error but a linear relationship between altitude and position uncertainty. Furthermore, we find that flight velocity defines how far along its transient or steady phase the system is able to go in a given time without affecting its convergence properties, while also decreasing the slope of the position drift in the steady phase. Contrary to the intuition that motion alongst the longer camera width helps reduce position error and uncertainty by

maximizing the length of SLAM feature tracks, we also uncover a linear relationship between position error and the FOV of the non- motion axis that results in faster convergence when flying alongst the height of a camera with more HFOV than VFOV. Lastly, we demonstrate that xVIO's unknown-depth initialization method is prone to divergence during vertical profiles and flights over discontinuous terrains, and subsequently evaluate 3 potential SLAM feature inverse-depth initialization alternatives, finding that using the median inverse-depth of the N closest tracked SLAM features offers the best results given sufficient filter iterations using an IEKF, while using an LRF range measurement is a more reliable approach otherwise.

Next, we analyze the impact of linearization on position error and uncertainty, finding an exponential decay relationship with camera image rate and a linear relationship with the number of tracked SLAM features, but no relationship with either the IMU rate or the number of IEKF iterations. We continue with an evaluation of the process noise, finding a linear relationship between inertial noise and both the position error and its uncertainty, and demonstrate that inflating the approximated accelerometer and gyroscope noises by a factor of 4 and 2 respectively to obtain the process noise provides the right balance in filter confidence to maximize accuracy whilst minimizing risks of divergence.

Now focusing on Visual Odometry, we start with an analysis of quantization error where we observe an exponential decay relationship between image resolution and both position error and uncertainty. Conversely, we find that tracking error has a radical relationship with position error and a cubic relationship with position uncertainty. However, we demonstrate that motion blur does not contribute to this error given our sensor suite and trajectory designs, and observe no significant improvement in performance from incorporating the camera distortion model to our visual measurement model.

Moving to Range-Visual Odometry, we find no range-dependent noise nor any significant biases in our SF30D LRF based on our field experiments. Assuming this weren't the case, we observe a linear relationship between LRF noise and both position error and uncertainty in simulation. Furthermore, while we demonstrate that horizontal violations of xVIO's zero camera-LRF translation assumption do

not affect performance, vertical offsets lead to errors in the velocity estimates which increase the position drift rate. Lastly, we find that while the local planarity assumption of the Delaunay triangles in xVIO's range update holds well over planar terrains, the range measurement uncertainty can rapidly grow over discontinuous terrain and become unrecoverable, leading subsequent range updates to be rejected as outliers by the Mahalanobis distance check for the remainder of the flight.

Finally, we evaluate xVIO's measurement noises. We find that a minimum of 6σ, where σ represents the true normalized pixel standard deviation, is required as visual measurement noise to prevent occurrences of filter divergence (<0.01% risk) in our large Monte Carlo simulations (M=300). Above this boundary, we find a linear relationship between visual measurement noise and both position error and uncertainty. Conversely, while we do not find the filter consistency to be sensitive to the range measurement noise, we observe an exponential decay relationship between the range measurement noise and both the position error and covariance.

4.2 Validation with high-altitude outdoor flights

We demonstrate xVIO's high-altitude capabilities in the Arroyo Secco environment next to JPL via a series of flights ranging up to 1km in length at 135m in height using our modified Sentinel UAV, achieving with RVIO less than 2% of final position error in all axes relative to the distance travelled, compared to less than 10% with VIO only.

5 Conclusion and future work

In conclusion, this work analyzed and discussed state-of-the-art high-altitude vision-based navigation, evaluated the high-altitude and all-terrain capabilities and limitations of the xVIO range-visual-inertial odometry framework, and demonstrated its high-altitude and long-range capabilities in the field.

Extending this work in the future, the NASA Jet Propulsion Laboratory will likely conduct outdoor experiments at higher altitudes, over longer distances and in terrains that better reflect Mars, using a more capable drone. Furthermore, while unfortunately unavailable for this research, the newly launched Pegasus aerial robotics simulator [107] and the upcoming Microsoft Project AirSim simulator are opening many new realms of possibilities for realistic simulations of aerial autonomy on the Martian surface. Areas of particular interest for future research include the evaluation of Multi-Image Super-Resolution (MISR) techniques to mitigate quantization error at high altitudes, and the evaluation of Visual Compass techniques to mitigate heading drift over long trajectories.

A IMU sensor model

Parameter	Definition	Unit
dt	Measurement frequency	Hz
σ_a	Discrete accel noise spectral density	$m/s^2/sqrt(Hz)$
σ_{ba}	Discrete accel bias random walk	$m/s^3/sqrt(Hz)$
σ_w	Discrete gyro noise spectral density	rad/s/sqrt(Hz)
σ_{bw}	Discrete gyro bias random walk	$rad/s^2/sqrt(Hz)$
T_s	G-sensitivity matrix	
T_a	Accel misalignment matrix	
T_g	Gyro misalignment matrix	
b_a	Accel bias	m/s
b_w	Gyro bias	rad/s
m_a	Accel measurement	m/s^2
m_w	Gyro measurement	rad/s^2
Φ	Random gaussian distributed number	
g	Gravity	m/s^2

Table A.1: IMU model parameters

Given the above parameters and the current real position p, orientation R, velocity v, acceleration a, and angular rate w of the IMU, we can write :

Accelerometer measurement model	Gyroscope measurement model
$b_a^t = b_a^{t-1} + dt \cdot \sigma_{ba} \cdot \Phi$	$b_w^t = b_w^{t-1} + dt \cdot \sigma_{bw} \cdot \Phi$
$a = T_a \cdot R \cdot (a - g)$	$w = (T_g \cdot w) + (T_s \cdot f_a)$
$m_a = a + b_a + \sigma_a \cdot \Phi$	$m_w = w + b_w + \sigma_w \cdot \Phi$

B FOV distortion model

From [75], we learn that given a camera distortion coefficient s, we can distort a 2D pixel $p_u(u_u, v_u)$ s.t :

$$p_d(u_d, v_d) = \frac{r_d}{r_u} \cdot \begin{bmatrix} u_u \\ v_u \end{bmatrix} \tag{B.1}$$

using the undistorted radius

$$r_u = \sqrt{u_u^2 + v_u^2} \tag{B.2}$$

and the distorted radius

$$r_d = \frac{tan^{-1}(2 \cdot r_u \cdot tan(\frac{s}{2})))}{s} \tag{B.3}$$

From this model, we can derive the distortion Jacobian

$$J_{d_{(2\cdot 2)}} \equiv \begin{bmatrix} \frac{\delta u_d}{u_u} & \frac{\delta u_d}{v_u} \\ \frac{\delta v_d}{u_u} & \frac{\delta v_d}{v_u} \end{bmatrix} \tag{B.4}$$

$$\begin{aligned} \frac{\delta u_d}{u_u} = & \frac{tan^{-1}(2 \cdot tan(\frac{s}{2})) \cdot r_u}{s \cdot r_u} - \frac{u_u^2 \cdot tan^{-1}(2 \cdot tan(\frac{s}{2})) \cdot r_u}{s \cdot r_u^2} \\ & + \frac{2 \cdot u_u^2 \cdot tan(\frac{s}{2})}{s \cdot (u_u^2 + v_u^2) \cdot (4 \cdot tan^2(\frac{s}{2}) \cdot (u_u^2 + v_u^2) + 1)} \end{aligned} \tag{B.5}$$

$$\begin{aligned} \frac{\delta u_d}{v_u} = \frac{\delta v_d}{u_u} = & \frac{2 \cdot u_u \cdot v_u \cdot tan(\frac{s}{2})}{s \cdot (u_u^2 + v_u^2) \cdot (4 \cdot tan^2(\frac{s}{2}) \cdot (u_u^2 + v_u^2) + 1} \\ & - \frac{u_u \cdot v_u \cdot tan^{-1}(2 \cdot tan(\frac{s}{2}))) \cdot r_u}{s \cdot r_u^2} \end{aligned} \tag{B.6}$$

$$\begin{aligned} \frac{\delta v_d}{v_u} = & \frac{tan^{-1}(2 \cdot tan(\frac{s}{2})) \cdot r_u}{s \cdot r_u} - \frac{v_u^2 \cdot tan^{-1}(2 \cdot tan(\frac{s}{2})) \cdot r_u}{s \cdot r_u^2} \\ & + \frac{2 \cdot v_u^2 \cdot tan(\frac{s}{2})}{s \cdot (u_u^2 + v_u^2) \cdot (4 \cdot tan^2(\frac{s}{2}) \cdot (u_u^2 + v_u^2) + 1)} \end{aligned} \tag{B.7}$$

C Reference parameter values

Parameter	Unit	Value
Distance travelled	m (x,y,z)	[300,0,0]
Terrain slope	degrees	0

Table C.1: Reference trajectory and terrain parameters

Parameter	Unit	Value
Initial position	m (x,y,z)	[0,0,100]
Initial velocity	m/s (x,y,z)	[10,0,0]
Initial attitude	degrees (x,y,z)	[0,0,0]
Initial gyro biases	$m/s^2/sqrt(Hz)$	0
Initial accel biases	rad/s/sqrt(Hz)	0
Initial position uncertainty	m	0
Initial velocity uncertainty	m/s	0.05
Initial attitude uncertainty	degrees	3
Initial accel bias uncertainty	m/s^2	0.3
Initial gyro bias uncertainty	deg/s	6
Gravity	m/s^2	[0.0, 0.0, -9.81]

Table C.2: Reference filter initialization parameters

Parameter	Unit	Value
Accel noise spectral density	$m/s^2/sqrt(Hz)$	0.0083
Accel bias random walk	$m/s^3/sqrt(Hz)$	0.00083
Gyro noise spectral density	rad/s/sqrt(Hz)	0.0013
Gyro bias random walk	$rad/s^2/sqrt(Hz)$	0.00013
Update frequency	Hz	210

Table C.3: Reference IMU parameters

Parameter	Unit	Value
Camera focal point	pixels (x,y)	[257.17, 254.75]
Camera principal point	pixels (x,y)	[354.04, 235.46]
Camera FOV distortion coefficient	ratio	0.93439
Camera resolution	pixels (x,y)	[640,480]
Camera-IMU translation	m (x,y,z)	[0,0,0]
Camera-IMU rotation	m (x,y,z)	[0,90,0]
Camera measurement noise	pixels	1
Update frequency	Hz	30

Table C.4: Reference camera parameters

Parameter	Unit	Value
LRF intersect point	pixels (x,y)	[353, 235]
LRF-Camera translation	m (x,y,z)	[0,0,0]
LRF-Camera rotation	m (x,y,z)	[0,0,0]
Range measurement noise	m	1
Update frequency	Hz	30

Table C.5: Reference range parameters

Parameter	Unit	Value
FAST detection treshold	pixels	20
Minimum feature count	features	400
Number of tiles	tiles	9
Max features per tile	features	200

Table C.6: Reference feature tracker parameters

Bibliography

[1] Theodore Tzanetos et al. "Ingenuity Mars Helicopter: From Technology Demonstration to Extraterrestrial Scout". In: *2022 IEEE Aerospace Conference (AERO)*. 2022, pp. 01–19. DOI: 10.1109/AERO53065.2022.9843428.

[2] Bob Balaram et al. "Mars Helicopter Technology Demonstrator". In: Jan. 2018. DOI: 10.2514/6.2018-0023.

[3] Benjamin T. Pipenberg et al. "Conceptual Design of a Mars Rotorcraft for Future Sample Fetch Missions". In: *2022 IEEE Aerospace Conference (AERO)*. 2022, pp. 01–14. DOI: 10.1109/AERO53065.2022.9843820.

[4] NASA. *Moon to Mars Architecture*. https://www.nasa.gov/moontomarsarchitecture/. [Online; accessed 18-October-2023]. 2023.

[5] Shannah Withrow et al. "Mars Science Helicopter Conceptual Design". In: Nov. 2020. DOI: 10.2514/6.2020-4029.

[6] Jonathan Bapst et al. "Mars Science Helicopter: Compelling Science Enabled by an Aerial Platform". In: *Bulletin of the AAS* 53 (Mar. 2021). DOI: 10.3847/25c2cfeb.a126aea0.

[7] Theodore Tzanetos et al. "Future of Mars Rotorcraft - Mars Science Helicopter". In: *2022 IEEE Aerospace Conference (AERO)*. 2022, pp. 1–16. DOI: 10.1109/AERO53065.2022.9843501.

[8] NASA. *Mars Atlas VALLES MARINERIS*. https://mars.nasa.gov/gallery/atlas/valles-marineris.html. [Online; accessed 5-June-2023].

[9] NASA JPL Mars Communications Team. *Mars Helicopter Flight Log*. https://mars.nasa.gov/technology/helicopter/. [Online; accessed 5-June-2023].

[10] Roland Brockers et al. "On-board Absolute Localization Based on Orbital Imagery for a Future Mars Science Helicopter". In: *2022 IEEE Aerospace Conference (AERO)*. 2022, pp. 1–11. DOI: 10.1109/AERO53065.2022.9843673.

[11] Jeff Delaune, David S. Bayard, and Roland Brockers. "Range-Visual-Inertial Odometry: Scale Observability Without Excitation". In: *IEEE Robotics and Automation Letters* 6.2 (2021), pp. 2421–2428. DOI: 10.1109/LRA.2021.3058918.

[12] NASA. *HiRISE - The University of Arizona*. https://www.uahirise.org/epo/about/. [Online; accessed 18-October-2023].

[13] NASA. *HiRISE*. https://mars.nasa.gov/mro/mission/instruments/hirise/. [Online; accessed 18-October-2023].

[14] ovt.com. *OV7251*. http://www.ovt.com/products/ov7251/. [Accessed 05-Jun-2023].

[15] Jeff Delaune, David S. Bayard, and Roland Brockers. *xVIO: A Range-Visual-Inertial Odometry Framework*. 2020. arXiv: 2010.06677 [cs.RO].

[16] Jeff Delaune et al. "Extended Navigation Capabilities for a Future Mars Science Helicopter Concept". In: *2020 IEEE Aerospace Conference*. 2020, pp. 1–10. DOI: 10.1109/AERO47225.2020.9172289.

[17] Jiaxin Hu et al. "1D-LRF Aided Visual-Inertial Odometry for High-Altitude MAV Flight". In: *2022 International Conference on Robotics and Automation (ICRA)*. 2022, pp. 5858–5864. DOI: 10.1109/ICRA46639.2022.9811757.

[18] Martin Scheiber et al. "Mid-Air Range-Visual-Inertial Estimator Initialization for Micro Air Vehicles". In: *2021 IEEE International Conference on Robotics and Automation (ICRA)*. 2021, pp. 7613–7619. DOI: 10.1109/ICRA48506.2021.9560913.

[19] Jeff Delaune et al. "Thermal-Inertial Odometry for Autonomous Flight Throughout the Night". In: *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. 2019, pp. 1122–1128. DOI: 10.1109/IROS40897.2019.8968238.

[20] Florian Mahlknecht et al. "Exploring Event Camera-based Odometry for Planetary Robots". In: *IEEE Robotics and Automation Letters (RA-L)* (2022).

[21] Franz Andert and Luis Mejias. "Improving monocular SLAM with altimeter hints for fixed-wing aircraft navigation and emergency landing". In: *2015 International Conference on Unmanned Aircraft Systems (ICUAS)*. 2015, pp. 1008–1016. DOI: https://doi.org/10.1109/ICUAS.2015.7152390.

[22] Michael Warren and Ben Upcroft. "High Altitude Stereo Visual Odometry". In: June 2013. DOI: https://doi.org/10.15607/RSS.2013.IX.003.

[23] M. Warren, P. Corke, and B. Upcroft. "Long-range stereo visual odometry for extended altitude flight of unmanned aerial vehicles". In: *The International Journal of Robotics Research* 35 (2015), pp. 381–403. DOI: https://doi.org/10.1177/0278364915581194.

[24] Anand George et al. "Visual-Inertial Odometry Using High Flying Altitude Drone Datasets". In: *Drones* 7.1 (2023). ISSN: 2504-446X. DOI: https://doi.org/10.3390/drones7010036. URL: https://www.mdpi.com/2504-446X/7/1/36.

[25] Caballero, F. and Merino, L. and Ferruz, J. et al. "Vision-Based Odometry and SLAM for Medium and High Altitude Flying UAVs." In: *Journal for Intellgen and Robotic Systems* 54 (2009), pp. 137–161. DOI: https://doi.org/10.1007/s10846-008-9257-y.

[26] Muhammad Yeasir Arafat, Muhammad Alam, and Sangman Moh. "Vision-Based Navigation Techniques for Unmanned Aerial Vehicles: Review and Challenges". In: *Drones* 7 (Jan. 2023), p. 89. DOI: 10.3390/drones7020089.

[27] Rokas Jurevičius and Virginijus Marcinkevičius. "A data set of aerial imagery from robotics simulator for map-based localization systems benchmark". In: *International Journal of Intelligent Unmanned Systems* 8.3 (2019), pp. 177–186. DOI: 10.1108/ijius-12-2018-0035. URL: https://doi.org/10.1108%2Fijius-12-2018-0035.

[28] Federal Aviation Administration (FAA). "Federal Aviation Regulation 91.119". In: *Federal Aviation Regulations (FARs)* (). [Online; accessed 22-Oct-2023].

[29] the European Commission. "REGULATION (EU) No 923/2012". In: *Official Journal of the European Union* (). [Online; accessed 22-Oct-2023].

[30] Jenny Nilsson Boij. "Localization of Combat Aircraft at High Altitude using Visual Odometry". https://www.diva-portal.org/smash/get/diva2:1677039/FULLTEXT01.pdf. MA thesis. Umea, Sweden: Umea University, 2022.

[31] Manjula Sharma et al. "Survey on Unmanned Aerial Vehicle for Mars Exploration: Deployment Use Case". In: *Drones* 6.1 (2022). ISSN: 2504-446X. DOI: 10.3390/drones6010004. URL: https://www.mdpi.com/2504-446X/6/1/4.

[32] I. Urzua, Rodrigo Munguía, and Antoni Grau. "Vision-based SLAM system for MAVs in GPS-denied environments". In: *International Journal of Micro Air Vehicles* 9 (June 2017), p. 175682931770532. DOI: 10.1177/1756829317705325.

[33] Mohammad O. A. Aqel et al. "Review of visual odometry: types, approaches, challenges, and applications". In: *SpringerPlus* 5 (2016). URL: https://api.semanticscholar.org/CorpusID:17005811.

[34] Ji Zhang and Sanjiv Singh. "INS Assisted Monocular Visual Odometry for Aerial Vehicles". In: *International Symposium on Field and Service Robotics*. 2013. URL: https://api.semanticscholar.org/CorpusID:13149697.

[35] C. Cadena et al. "Past, present, and future of simultaneous localization and mapping: Toward the robust-perception age". In: *IEEE Transactions on Robotics* 32.6 (Dec. 2016), pp. 1309–1332.

[36] David Bayard et al. "Vision-Based Navigation for the NASA Mars Helicopter". In: Jan. 2019. DOI: 10.2514/6.2019-1411.

[37] Tianbo Liu and Shaojie Shen. "High altitude monocular visual-inertial state estimation: Initialization and sensor fusion". In: *2017 IEEE International Conference on Robotics and Automation (ICRA)*. 2017, pp. 4544–4551. DOI: 10.1109/ICRA.2017.7989528.

[38] Andrew Lambert et al. "Visual odometry aided by a sun sensor and inclinometer". In: *2011 Aerospace Conference*. 2011, pp. 1–14. DOI: 10.1109/AERO.2011.5747268.

[39] A. Trebi-Ollennu et al. "Design and analysis of a sun sensor for planetary rover absolute heading detection". In: *IEEE Transactions on Robotics and Automation* 17.6 (2001), pp. 939–947. DOI: 10.1109/70.976028.

[40] Teddy Tzanetos and Ingenuity Team Lead at NASA's Jet Propulsion Laboratory. *STATUS UPDATES — May 27, 2022 : Ingenuity Adapts for Mars Winter Operations*. https://mars.nasa.gov/technology/helicopter/status/382/ingenuity-adapts-for-mars-winter-operations/. [Online; accessed 5-October-2023].

[41] Vincenzo Angelino et al. "High altitude UAV navigation using IMU, GPS and camera". In: *Proceedings of the 16th International Conference on Information Fusion*. 2013, pp. 647–654.

[42] Chaofan Zhang et al. "VINS-MKF: A Tightly-Coupled Multi-Keyframe Visual-Inertial Odometry for Accurate and Robust State Estimation". In: *Sensors* 18 (Nov. 2018), p. 4036. DOI: 10.3390/s18114036.

[43] Sturm, J. and Engelhard, N. and Endres, F. and Burgard, W. and Cremers, D. "A benchmark for the evaluation of RGB-D SLAM systems". In: *Proceedings of the 2012 IEEE/RSJ International Conference on Intelligent Robots and Systems*. 2012, pp. 573–580.

[44] D. Prokhorov, D. Zhukov, O. Barinova, K. Anton, and A. Vorontsova. "Measuring robustness of visual slam". In: *16th International Conference on Machine Vision Applications (MVA)* (2019), pp. 1–6.

[45] Amey Kasar. *Benchmarking and Comparing Popular Visual SLAM Algorithms*. Nov. 2018.

[46] Davide Scaramuzza and Friedrich Fraundorfer. "Visual Odometry [Tutorial]". In: *IEEE Robotics and Automation Magazine* 18.4 (2011), pp. 80–92. DOI: 10.1109/MRA.2011.943233.

[47] Renhan Wu and Yuzhuo Gao. "Research on Underwater Complex Scene SLAM Algorithm Based on Image Enhancement". In: *Sensors* 22.21 (2022). ISSN: 1424-8220. DOI: 10.3390/s22218517. URL: https://www.mdpi.com/1424-8220/22/21/8517.

[48] M. Amiri Atashgah and S. Malaek. "An integrated virtual environment for feasibility studies and implementation of aerial monoslam". In: *Virtual Reality* 16 (2012), pp. 215–232. DOI: https://doi.org/10.1007/s10055-011-0197-7.

[49] Rudolf E. Kalman and Richard S. Bucy. "New Results in Linear Filtering and Prediction Theory". In: *Journal of Basic Engineering* 83 (1961), pp. 95–108. URL: https://api.semanticscholar.org/CorpusID:8141345.

[50] Edward Rosten and Tom Drummond. "Machine Learning for High-Speed Corner Detection". In: *European Conference on Computer Vision*. 2006. URL: https://api.semanticscholar.org/CorpusID:1388140.

[51] J.-Y. Bouguet. "Pyramidal implementation of the lucas kanade feature tracker". In: 1999. URL: https://api.semanticscholar.org/CorpusID:9350588.

[52] Jianbo Shi and Tomasi. "Good features to track". In: *1994 Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*. 1994, pp. 593–600. DOI: 10.1109/CVPR.1994.323794.

[53] Mingyang Li and Anastasios Mourikis. "Optimization-Based Estimator Design for Vision-Aided Inertial Navigation". In: July 2012. DOI: 10.15607/RSS.2012.VIII.031.

[54] B Delaune. "Sur la sphere vide. A la memoire de Georges Voronoi". In: *Bulletin de l'Academie des Sciences de l'URSS* (6 1934), pp. 793–800.

[55] Javier Civera, Andrew J. Davison, and J. M. MartÍnez Montiel. "Inverse Depth Parametrization for Monocular SLAM". In: *IEEE Transactions on Robotics* 24.5 (2008), pp. 932–945. DOI: 10.1109/TRO.2008.2003276.

[56] Brandon Wagstaff and Jonathan Kelly. "Self-Supervised Scale Recovery for Monocular Depth and Egomotion Estimation". In: *2021 IEEE/RSJ International Conference on Intel-*

ligent Robots and Systems (IROS). IEEE, 2021. DOI: 10.1109/iros51168.2021.9635938. URL: https://doi.org/10.1109%2Firos51168.2021.9635938.

[57] Rudolf E. Kalman. "A New Approach to Linear Filtering and Prediction Problems". In: *Transactions of the ASME–Journal of Basic Engineering* 82.Series D (1960), pp. 35–45.

[58] Alex Becker. *Kalman filter from the ground up*. 2nd. Becker, Alex, 2023. ISBN: 978-965-93120-1-6.

[59] Guoquan P. Huang, Anastasios I. Mourikis, and Stergios I. Roumeliotis. "Analysis and improvement of the consistency of extended Kalman filter based SLAM". In: *2008 IEEE International Conference on Robotics and Automation*. 2008, pp. 473–479. DOI: 10.1109/ROBOT.2008.4543252.

[60] S. Julier. "The stability of covariance inflation methods for SLAM". In: *Proceedings 2003 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2003) (Cat. No.03CH37453)*. Vol. 3. 2003, 2749–2754 vol.3. DOI: 10.1109/IROS.2003.1249286.

[61] T. C. Dong-Si and A. I. Mourikis. "Estimator initialization in vision-aided inertial navigation with unknown camera-IMU calibration". In: *IEEE International Conference on Intelligent Robots and Systems* (2012), pp. 1064–1071.

[62] A. Martinelli. "Closed-Form Solution of Visual-Inertial Structure from Motion". In: *International Journal of Computer Vision* 16.2 (2014), pp. 138–152.

[63] Tianbo Liu and Shaojie Shen. "Spline-Based Initialization of Monocular Visual–Inertial State Estimators at High Altitude". In: *IEEE Robotics and Automation Letters* 2.4 (2017), pp. 2224–2231. DOI: 10.1109/LRA.2017.2724770.

[64] You Li et al. "IMU/Magnetometer/Barometer/Mass-Flow Sensor Integrated Indoor Quadrotor UAV Localization with Robust Velocity Updates". In: *Remote Sensing* 11 (Apr. 2019), p. 838. DOI: 10.3390/rs11070838.

[65] Ying Liu et al. "A Visual Compass Based on Point and Line Features for UAV High-Altitude Orientation Estimation". In: *Remote Sensing* 14.6 (2022). ISSN: 2072-4292. DOI: 10.3390/rs14061430. URL: https://www.mdpi.com/2072-4292/14/6/1430.

[66] Gianpaolo Conte and Patrick Doherty. "Vision-Based Unmanned Aerial Vehicle Navigation Using Geo-Referenced Information". In: *EURASIP Journal on Advances in Signal Processing* (2009). DOI: https://doi.org/10.1155/2009/387308.

[67] lumenera. *The Challenge of Aerial Imaging: Achieving a Clear and Sharp Image Teledyne Lumenera*. https://www.lumenera.com/blog/the-challenge-of-aerial-imaging-achieving-a-clear-and-sharp-image. [Accessed 05-Jun-2023]. 2018.

[68] DJI. *Ground Sample Distance*. https://enterprise-insights.dji.com/blog/ground-sample-distance. [Accessed 05-Jun-2023]. 2022.

[69] Magnus Offermans. "Map-based localization for micro air vehicles autonomous navigation". https://www.marshallplan.at/images/All-Papers/MP-2019/Magnus_Offermanns_Marshallplan_Field+Report.pdf. MA thesis. Klagenfurt, Austria: Alpen-Adria-Universität Klagenfurt, 2020.

[70] Qiang Zhao and Limei Peng. "High-altitude Multi-object Detection and Tracking based on Drone Videos". In: *Journal of Networking and Network Applications* 2 (2022), pp. 36–42. DOI: 10.33969/J-NaNA.2022.020103.

[71] Marcus Märtens et al. *Super-Resolution of PROBA-V Images Using Convolutional Neural Networks*. 2019. arXiv: 1907.01821 [cs.CV].

[72] Woyeong Jeong, Jiyoun Moon, and Beom Lee. "Error Improvement in Visual Odometry Using Super-resolution". In: *International Journal of Control, Automation and Systems* 18 (Dec. 2019). DOI: 10.1007/s12555-019-0256-5.

[73] Bruce Lucas and Takeo Kanade. "An Iterative Image Registration Technique with an Application to Stereo Vision (IJCAI)". In: vol. 81. Apr. 1981.

[74] David S. Bayard et al. "Vision-Based Navigation for the NASA Mars Helicopter". In: *AIAA Scitech 2019 Forum*. 2019. DOI: 10.2514/6.2019-1411. eprint: https://arc.aiaa.org/doi/pdf/10.2514/6.2019-1411. URL: https://arc.aiaa.org/doi/abs/10.2514/6.2019-1411.

[75] Frederic Devernay and Olivier D. Faugeras. "Straight lines have to be straight". In: *Machine Vision and Applications* 13 (2001), pp. 14–24.

[76] A Terui. "Revisiting scale invariance and scaling in ecology: River fractals as an example". In: *Population Ecology* (2023), pp. 1–7. DOI: https://doi.org/10.1002/1438-390X.12163.

[77] Martial Sanfourche et al. "Perception for UAV: Vision-Based Navigation and Environment Modeling." In: 2012. URL: https://api.semanticscholar.org/CorpusID:17651900.

[78] Christopher Mei and Ian Reid. "Modeling and generating complex motion blur for real-time tracking". In: *2008 IEEE Conference on Computer Vision and Pattern Recognition*. 2008, pp. 1–8. DOI: 10.1109/CVPR.2008.4587535.

[79] Tobias Senst, Volker Eiselein, and Thomas Sikora. "Robust Local Optical Flow for Feature Tracking". In: *IEEE Transactions on Circuits and Systems for Video Technology* 22 (Sept. 2012). DOI: 10.1109/TCSVT.2012.2202070.

[80] Michael Warren et al. "Large Scale Monocular Vision-Only Mapping from a Fixed-Wing sUAS". In: vol. 92. 2014, pp. 495–509. DOI: https://doi.org/10.1007/978-3-642-40686-7_33.

[81] Hauke Malte Strasdat, José M. M. Montiel, and Andrew J. Davison. "Scale Drift-Aware Large Scale Monocular SLAM". In: *Robotics: Science and Systems*. 2010. URL: https://api.semanticscholar.org/CorpusID:9307872.

[82] Martin Scheiber. "Range-Visual-Inertial Initialization For Micro Aerial Vehicles". https://netlibrary.aau.at/obvuklhs/download/pdf/7034133?originalFilename=true. MA thesis. Klagenfurt, Austria: Alpen-Adria-Universität Klagenfurt, 2020.

[83] Martin Wudenka et al. "Towards Robust Monocular Visual Odometry for Flying Robots on Planetary Missions". In: *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)* (2021), pp. 8737–8744.

[84] A. Cherian et al. "Autonomous altitude estimation of a UAV using a single onboard camera". In: *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. 2009, pp. 3900–3905.

[85] Haolong Li and Joerg Stueckler. *Observability Analysis of Visual-Inertial Odometry with Online Calibration of Velocity-Control Based Kinematic Motion Models*. 2022. arXiv: 2204.06651 [cs.RO].

[86] Dimitrios G. Kottas, Kejian J. Wu, and Stergios I. Roumeliotis. "Detecting and dealing with hovering maneuvers in vision-aided inertial navigation systems". In: *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems*. 2013, pp. 3172–3179. DOI: 10.1109/IROS.2013.6696807.

[87] Wenhui Xu et al. "A New Orientation Estimation Method Based on Rotation Invariant Gradient for Feature Points". In: *IEEE Geoscience and Remote Sensing Letters* 18.5 (2021), pp. 791–795. DOI: 10.1109/LGRS.2020.2985358.

[88] David S. Bayard. "An Overview of the Mars Helicopter Vision Based Navigation System". In: *International Conference on Intelligent Robotic Systems (IROS)*. 2021.

[89] Daniel Meier et al. "Detection and characterization of moving objects with aerial vehicles using inertial-optical flow". In: *2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. 2015, pp. 2473–2480. DOI: 10.1109/IROS.2015.7353713.

[90] Valentin Peretroukhin, Lee Clement, and Jonathan Kelly. "Reducing drift in visual odometry by inferring sun direction using a Bayesian Convolutional Neural Network". In: *2017 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2017. DOI: 10.1109/icra.2017.7989235. URL: https://doi.org/10.1109%2Ficra.2017.7989235.

[91] Valentin Peretroukhin, Lee Clement, and Jonathan Kelly. "Inferring sun direction to improve visual odometry: A deep learning approach". In: *The International Journal of Robotics Research* 37.9 (2018), pp. 996–1016. DOI: 10.1177/0278364917749732. eprint: https://doi.org/10.1177/0278364917749732. URL: https://doi.org/10.1177/0278364917749732.

[92] Lee Clement, Valentin Peretroukhin, and Jonathan Kelly. "Improving the Accuracy of Stereo Visual Odometry Using Visual Illumination Estimation". In: *2016 International Symposium on Experimental Robotics*. Ed. by Dana Kulić et al. Cham: Springer International Publishing, 2017, pp. 409–419.

[93] Ying Liu et al. "Uncalibrated downward-looking UAV visual compass based on clustered point features". In: *Science China Information Sciences* 62 (2019). DOI: 10.1007/s11432-018-9748-1.

[94] Wenhui Xu et al. "A New Orientation Estimation Method Based on Rotation Invariant Gradient for Feature Points". In: *IEEE Geoscience and Remote Sensing Letters* 18.5 (2021), pp. 791–795. DOI: 10.1109/LGRS.2020.2985358.

[95] J.L. Bufton. "Laser altimetry measurements from aircraft and spacecraft". In: *Proceedings of the IEEE* 77.3 (1989), pp. 463–477. DOI: 10.1109/5.24131.

[96] Fang Chen and Lingling Luo. In: *Open Physics* 20.1 (2022), pp. 1194–1202. DOI: doi:10.1515/phys-2022-0210. URL: https://doi.org/10.1515/phys-2022-0210.

[97] Andreas Eisele. "Millimeter-Precision Laser Rangefinder Using a Low-Cost Photon Counter". PhD thesis. 2014. 267 pp. ISBN: 978-3-7315-0152-7. DOI: 10.5445/KSP/1000037714.

[98] Cezary Kownacki. "Design of an adaptive Kalman filter to eliminate measurement faults of a laser rangefinder used in the UAV system". In: *Aerospace Science and Technology* 41 (2015), pp. 81–89.

[99] S. Bell. "Experimental Design". In: *International Encyclopedia of Human Geography*. Ed. by Rob Kitchin and Nigel Thrift. Oxford: Elsevier, 2009, pp. 672–675. ISBN: 978-0-08-044910-4. DOI: https://doi.org/10.1016/B978-008044910-4.00431-4. URL: https://www.sciencedirect.com/science/article/pii/B9780080449104004314.

[100] Shital Shah et al. "AirSim: High-Fidelity Visual and Physical Simulation for Autonomous Vehicles". In: 2018, pp. 621–635. ISBN: 978-3-319-67360-8. DOI: 10.1007/978-3-319-67361-5_40.

[101] Yunlong Song et al. "Flightmare: a flexible quadrotor simulator". In: Nov. 2020.

[102] Alessandro Fornasier et al. "VINSEval: Evaluation Framework for Unified Testing of Consistency and Robustness of Visual-Inertial Navigation System Algorithms". In: *2021 IEEE International Conference on Robotics and Automation (ICRA)*. 2021, pp. 13754–13760. DOI: 10.1109/ICRA48506.2021.9561382.

[103] N. Koenig and A. Howard. "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)*. Vol. 3. 2004, 2149–2154 vol.3. DOI: 10.1109/IROS.2004.1389727.

[104] Open Robotics. *Gazebo*. https://gazebosim.org/home. [Accessed 05-Jun-2023].

[105] Akash Patel et al. "Towards energy efficient autonomous exploration of Mars lava tube with a Martian coaxial quadrotor". In: *Advances in Space Research* 71.9 (2023). Application of Artificial Intelligence in Tracking Control and Synchronization of Spacecraft, pp. 3837–3854. ISSN: 0273-1177. DOI: https://doi.org/10.1016/j.asr.2022.11.014. URL: https://www.sciencedirect.com/science/article/pii/S0273117722010365.

[106] Joern Rehder et al. "Extending kalibr: Calibrating the extrinsics of multiple IMUs and of individual axes". In: *2016 IEEE International Conference on Robotics and Automation (ICRA)*. 2016, pp. 4304–4311. DOI: 10.1109/ICRA.2016.7487628.

[107] Marcelo Jacinto et al. *Pegasus Simulator: An Isaac Sim Framework for Multiple Aerial Vehicles Simulation*. 2023. arXiv: 2307.05263 [cs.RO].