ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

RSL Robotic Systems Lab
Prof. Dr. Marco Hutter

Master Thesis

Vision-based safe proximity operation of a future Mars rotorcraft

Autumn Term 2024

Supervised by:
Roland Brockers
Robert Hewitt
Marco Hutter

Author:
Lasse Fierz

Contents

List of Acronyms

- **UAV:** Unmanned Aerial Vehicle
- **SFM:** Structure From Motion
- **LSD:** Landing Site Detection
- **LS:** Landing Site
- **BA:** Bundle Adjustment
- **DEM:** Dense Elevation Map
- **OMG:** Optimal Mixture of Gaussian
- **LOD:** Level Of Detail
- **HiRISE:** High Resolution Imaging Science Experiment (High Resolution Satellite Imagery)
- **LRF:** Laser Range Finder
- **GT:** Ground Truth
- **LSM:** Landing Site Manager

Abstract

An autonomous rotorcraft literally stands or falls on it's reliable landing capabilities. When that same rotorcraft is on Mars, this procedure cannot fail even once. The LORNA (Long Range Navigation) project tackles this problem by introducing a Landing Site Detection (LSD) mechanism which aggregates Structure From Motion (SFM) point clouds into a multi resolution depth map and performs landing site segmentation on the collected depth information. In this master's thesis we incorporated this landing site detection pipeline into an autonomous framework and implemented a behavior tree based landing mechanism to safely and efficiently select, verify and discard detected landing sites. Furthermore the pipeline was enhanced using a stereo camera depth input alternative to SFM for lower altitudes to remove the necessity of lateral motion in order to perceive depth. The software was tested extensively in a gazebo simulation on different synthetic as well as recorded environments and different behaviors were considered and analyzed throughout various Monte Carlo iterations. The contributions in this work aim at enabling future mars rotorcrafts to autonomously and reliably land at safe locations thus enabling a more daring aerial exploration of the red planet.

Chapter 1

Introduction

With the Ingenuity rotorcraft's lifecycle coming to an end the question about future mars rotorcrafts and their capabilities draws ever closer.
For future large distance missions NASA is conceptualizing a Mars Science Helicopter (MSH) project. The aspirations for such a rotorcraft are on one hand to cover farther distances with accurate state estimation and on the other to land safely, autonomously and reliably in previously unknown terrain. The LOng Range NAvigation (LORNA) project that I have been involved with is working on a concept to tackle the aforementioned points while dealing with the constraints that rotorcraft missions on Mars provide us with. These are namely a limitation on the size and weight of the drone, a constraint on compuational power due to the deployment on limited embedded processors and lastly a delay in communication which makes adaptive remote control from Earth impossible.
On a high level the LORNA project consists of three parts.

1. xVIO: State Estimator - Merges camera images, IMU measurements and laser range finder information using an extended Kalman filter.
2. Landing Site Detection Pipeline - Uses structure from motion to aggregate point clouds and detect landing sites on the gathered data.
3. Autonomous Framework - Handles all high level flight behaviors and represents the interface between the flight controller, mission plan, landing site detector, system's healthguard and more.

In this thesis I put emphasis on the ladder two topics working with ground truth poses from the simulation environment.
Flying a rotorcraft on Mars is of course no new endeavor to NASA as the 71 successful flights performed by Ingenuity demonstrate. In contrast to Ingenuity however this work thrives towards the fully integrated usage of a landing site detection pipeline in an autonomous framework. This allows for a safe, reliable and most importantly autonomous landing procedure lifting the heavy safety constraints on the mission flown and reducing pre flight overhead considerably.

1.1 Objective

Concretly the endeavour in this thesis was to create a front to back landing mechanism that combines the existing vision based landing site detection procedure with the autonomous framework. In order to accomplish this, both the landing site detection mechanism as well as the autonomy had to be altered. Last but not least given that the structure from motion depth generation methodology depends on

lateral movement, which is less desirable for a drone navigating at low altitudes in unfamiliar surroundings, the utilization of a stereo camera presents a viable solution to attain real-time depth perception without necessitating lateral displacement.
The conclusive high level objective of the LORNA project is to autonomously fly a high altitude long distance science mission, using a map based localization enhanced state estimator. Whilst flying, landing sites are acquired and simultaneously ordered according to the respective quality. Initiating the adaptive autonomous landing sequence, landing sites can then be chosen and verified at low altitudes using a stereo camera. In case of successful verification the rotorcraft can land at the selected location.

Chapter 2

Related Work

With Ingenuity[1] NASA has already sent a drone to Mars. It was the first ever rotorcraft on another planet and instead of the initially planned 5 flights it completed an outstanding 72 flights in its lifecycle. However because it had no landing site detection capabilities every landing spot had to be determined in advance based on HiRISE images or rover footage. Therefore flight missions had to be planned with a large safety margin and considerable overhead.
The Long Range Navigation (LORNA) project that I have been involved with is part of NASA's Mars Science Helicopter mission which is the planned future rotorcraft mission taking on large distance traversal and scientific exploration. LORNA tries to tackle both the aforementioned reliable safe landing as well as large distance traversal in a three-step approach:

- State Estimation [2, 3]
- Landing Site Acquisition [4]
- Autonomous Framework [5]

In this thesis I expanded on these building blocks, connecting them in order to create a wholistic landing pipeline.

2.1 State Estimation

As this thesis was only scarcely linked to the state estimation of the project I will only superficially scratch that introduction.
Similar to the Ingenuity mission LORNA plans on using a Visual-Inertial Odometry state estimator called xVIO [2]. In contrast to the state estimator on Ingenuity, xVIO's modular approach allows it to use additional sensors like a laser range finder.
The xVIO state estimator receives the sensor outputs (Image, IMU, Laser Range Finder...) and using an Extended Kahlamn Filter, it creates a pose estimate out of it which is passed to the flight controller and the visual landing site acquisition nodes (SFM, LSD).

2.1.1 xMBL

Drift accumulated over time using this local state estimator can be reduced using a Map Based Localization(MBL) algorithm [3]. Hereby images from the drone are matched against a satellite map in order to refine the pose from VIO.

2.2 Landing Site Acquisition

A core part of the LORNA Mars Science Helicopter proposal is the landing site acquisition. As the terrain on Mars in high definition is mostly unknown and the landing procedure is the most dangerous part out of any rotorcraft mission it is imperative for a drone to perceive the elevation of the terrain beneath it and segment good landing sites on it.
The LORNA project tackles this in a two step approach:

- Depth Generation from Structure from Motion (SFM) [6]
- Landing Site Detection (LSD) [7, 8]

2.2.1 Structure from Motion (SFM)

The Structure from Motion algorithm [6] uses a keyframe based approach to create a point cloud from incoming monocular camera images. It receives images and their image poses respectively from the vision based state estimator. (2.1)
As it is supplied with monocular images from a nadir pointed camera it creates point clouds using lateral motion in between images. This allows for an adaptive baseline enabling depth perception at high altitudes. However at low altitudes lateral motion in unknown terrain poses a significant risk for a rotorcraft. More on this in section 5.1.

2.2.2 Landing Site Detection (LSD)

The Landing Site Detection algorithm is comprised of two steps:

- Depth Aggregation
- Landing Site Segmentation

Depth measurements are accumulated in a multi-resolution depth map. Initially as described in Schoppmann et al. [7] the different layers represented an absolute base layer and relative residual layers in the higher resolution representations.
In Proenca et al. [8] this was altered to only consider absolute measurements at each resolution layer.
However because of the initial implementation in [7], each layer considers the same map size.

2.3 Autonomous Framework

The autonomy [5] used by LORNA was develeped within the project itself.
It consists of a state machine which handles all fundamental behavior states and executes the respective node within a state. These nodes can be as simple as a unique standalone action. More complicated nodes handle sophisticated procedures using adaptive behavior trees.
In addition to the behavior exerted during nominal procedure the autonomy also handles system health events and hardware failure occurrences.
The autonomy embodies the interface between the flight controller, landing site detector, healthguard and more. It also contains an individual mission planner making the usage of an additional ground station obsolete.

Chapter 3

My Contribution

In short my contribution in this thesis can be summarized as follows:

3.1 Interface Autonomy-LSD

The autonomy was altered in a way to receive the detected landing sites and order them according to an adequate heuristic to determine the best landing site at any given time.

3.2 Landing Site Detection Output

The landing site detection output initially only consisted of the location of the landing site. This output was enhanced to consider many more characteristics in order for the autonomy to make an informed decision with regards to what spot to select.

3.3 Stereo Camera Depth Alternative

Implemented a stereo camera in the simulated drone model in order to get stereo sensor images. Created a stereo camera depth node as an alternative to SFM to supply the landing site detection algorithm with a point cloud at low altitudes without the need for lateral motion.

3.4 Additional

Other contributions indirectly connected to the project itself but necessary for the implementation:

3.4.1 Simulation Setup

As just recently the switch was made to Gazebo Garden the entire visual pipeline (SFM + LSD) had never run with this simulation environment before. Therefore I implemented the changes necessary to run the landing site detection procedure on the Gazebo sensor input. Additionally whilst implementing the stereo camera and attempting to put in place a simulated depth camera for ground truth it became apparent that the Gazebo depth camera implementation is incorrect neglecting the set intrinsic parameters. Altering Gazebo's source code the implementation could be fixed.

3.4.2 Deployment of LSD pipeline onto Embedded System

The entire software stack of this project has to run on an embedded system on the future rotorcraft. Currently the used processor is modalAI's voxl2. Both the structure from motion as well as the landing site detection software did not run out of the box having an incompatible dependency handling with the voxl's AARCH architecture. Resolving these issues I was able to run the landing site detection pipeline with the structure from motion depth supply on the voxl2 using a collected rosbag of images and IMU poses.

Chapter 4

Setup

Hereafter is an image depicting the high level structure of the LORNA project.

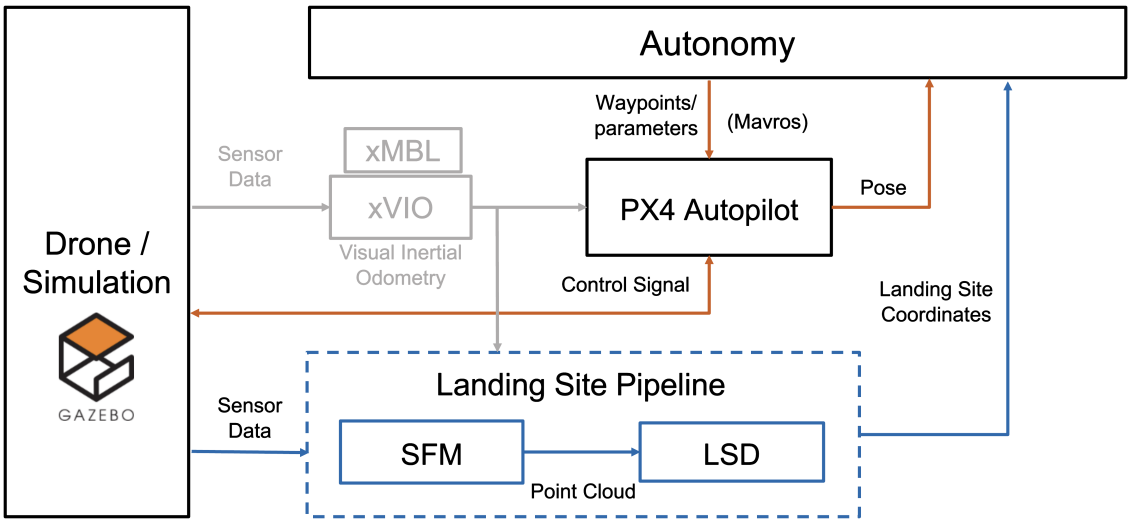

Figure 4.1: LORNA Project Setup

As this thesis revolved around the combination of existing software instances, it is essential to display the individual parts comprising the LORNA project in more detail.

4.1 Simulation

Despite being able to deploy the landing site detection pipeline onto the voxl2 processor the majority of this thesis was done using a Gazebo Garden simulation of the drone.

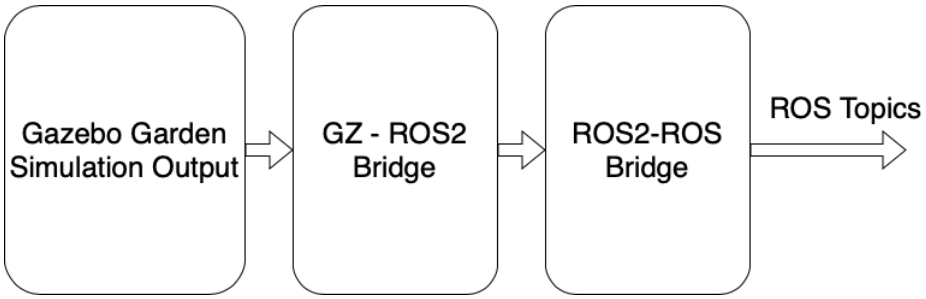

Figure 4.2: Gazebo ROS Bridge

As the entire software stack of the LORNA project is dependent on ROS instead of ROS2 a bridge was used to convert the sensor information from Gazebo to ROS2 and from ROS2 to ROS.

4.2 Landing Site Acquisition Pipeline

The landing site acquisition pipeline consists of two nodes. A structure from motion node [6] which creates a point cloud using a keyframe based stereo approach on monocular images and a landing site detector node [7, 8] which aggregates the depth measurements into a rolling buffer based multi-resolution depth map and segments landing sites on the created DEM. The found landing sites are then supplied to the autonomy.

4.2.1 Structure From Motion (SFM)

SFM - Bundle Adjustment

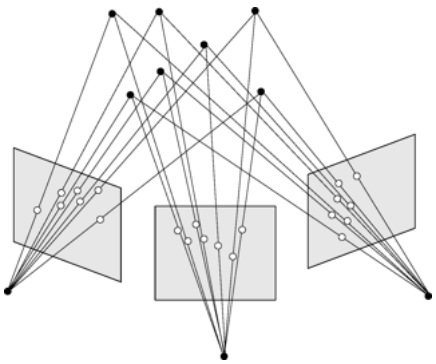

Figure 4.3: Bundle Adjustment Procedure

In a first step the keyframes are filled with the incoming images and their respective camera pose information. Once the keyframes are filled, each iteration the input as well as all the keyframe poses are refined using a Bundle Adjustment algorithm.

SFM - Stereo Depth

The keyframes and their refined poses are then compared to the new incoming image with regards to image overlap and feature retention. Chosing the most adequate keyframe and the incoming frame, one can create a depth image. The depth image is then converted into a point cloud and packaged together with the respective poses of the images. This allows not only to correctly locate the points in a global frame but also to derive the baseline with which that point cloud was created.

SFM - Keyframe Updates

At the end of an iteration the keyframes are updated based on the aforementioned characteristics of image overlap and feature retention quality.

4.2.2 Landing Site Detection (LSD)

LSD - Depth Aggregation

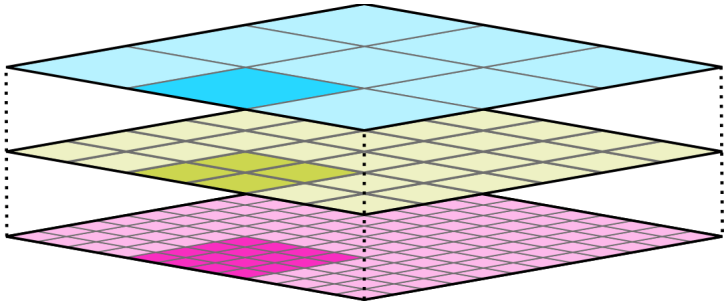

Figure 4.4: Multi Resolution Depth Map

The foundation of the landing site detection mechanism is a rolling buffer based multi resolution depth map as indicated in fig. 4.4. Each base layer cell is represented with 4 cells at a higher resolution layer.
Each point cloud input iteration, the measurements are placed in the respective cells based on the level of detail of the perceived points. In a subsequent step the measurements are pooled up and down the resolution layers in order to make the DEM more consistent and interpolate missing values.
Each cell in this dense elevation map (DEM) is comprised of an optimal mixture of gaussian (OMG) state as described in Proenca et al. [8].
Its update step looks like this:

$$S_t = S_{t-1} + \sigma_{x_t}^{-2} \tag{4.1}$$

$$\mu_t = \frac{1}{S_t}\left(S_{t-1}\mu_{t-1} + \frac{x_t}{\sigma_{x_t}^2}\right) \tag{4.2}$$

$$\sigma_t^2 = \frac{1}{S_t}\left(S_{t-1}(\sigma_{t-1}^2 + \mu t - 1^2) + \frac{x_t^2}{\sigma {x_t}^2} + 1\right) - \mu_t^2 \tag{4.3}$$

Where μ_t is the cell's new mean value, σ_t^2 is the cell's variance and S_t defines an auxilliary variable to keep track of all past variances within a single scalar.
Thus similar to Kalman filters, the OMG cells' uncertainties decline over time as more measurements are entered. Because of this the DEM's terrain estimate converges over time.

LSD - Hazard Segmentation

On the created depth map landing sites can then be detected. This is done using a roughness and slope assessment of the perceived terrain. Roughness defines the maximum absolute altitude difference around a cell in a certain resolution layer and slope is determined by fitting a plane to the vicinity of a considered point.
If the roughness and slope values lie within the acceptance threshold, the spot is recognized as a landing site and marked as such in a binary landing map.

Figure 4.5: Binary Landing Site Map

LSD - Landing Site Selection

After applying a distance transform on the landing site map and performing non-maximum suppression on the landing site sizes, the biggest landing sites are found. Their positions are then refined one last time using a mean shift algorithm that considers roughness, uncertainty and size once again.

Figure 4.6: Binary Landing Site Map after Non Maximum Suppression

LSD - Debug Images

Figure 4.7: Gazebo Simulation Reference

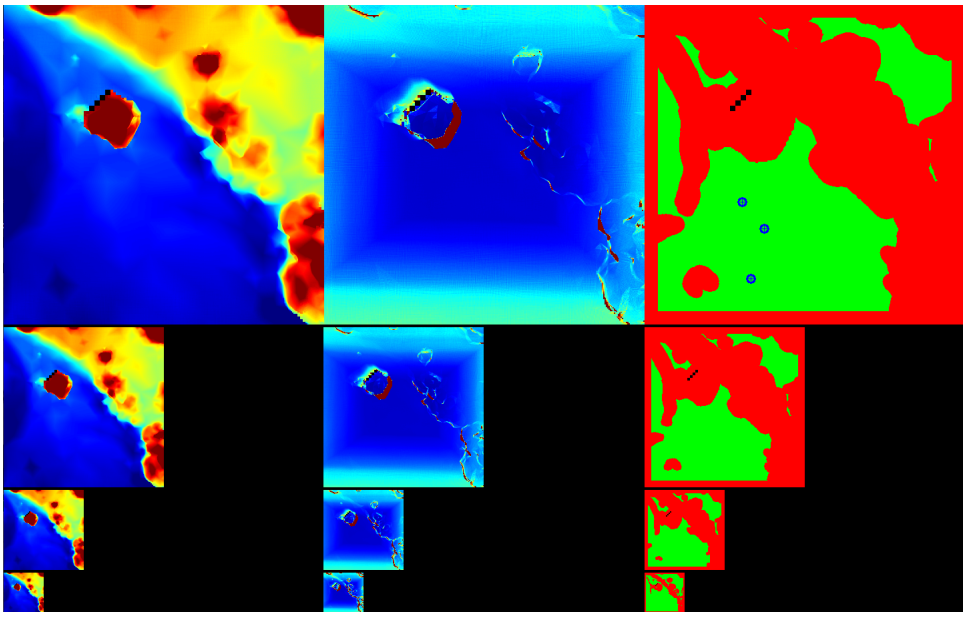

Figure 4.8: LSD Debug Image - Left: DEM, Middle: Uncertainties, Right: LS Map

The landing site detection debug image is a good comprehensive visualization of the landing spot detection procedure.
On the left one can see the multi-resolution map displaying the same terrain area in different resolutions. Red pixels are closer, blue further away.

In the middle one can see the uncertainties of the detected points. For simplification purposes the variance of the detected points is simply the associated stereo depth error.
On the right is the above mentioned binary landing site map. Green indicates valid landing sites, and the blue crosses indicate the chosen non-max suppressed and mean shifted landing sites.

4.3 Autonomy

The autonomous framework was developed within the LORNA project. It is the overarching instance governing all the necessary behaviors and constituting the interface between all the different nodes of the process. It is connected to the flight controller through the Mavros wrapper of the Mavlink protocol. With this connection it can send waypoints and parameter updates to the flight controller. In addition to the in fig. 4.1 shown connections it also communicates with a healthguard node keeping track of the systems health state and alerting in case of anomalies.

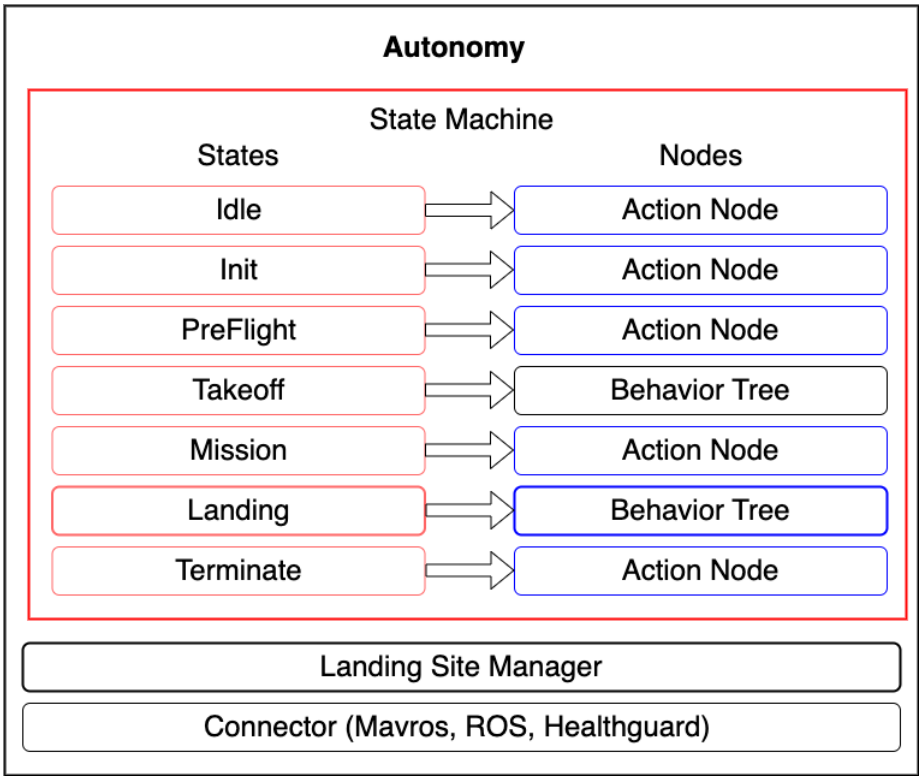

Figure 4.9: Simplified Structure of the Autonomy

The core of the autonomy is the state machine as depicted in fig. 4.9. In each state the respective node is executed which most often is a single action node. For more complicated procedures a behavior tree is used. This is the case for the takeoff as well as landing nodes. As indicated in bold, the landing node is the most crucial node in this work as this is where the landing behavior using the landing sites is executed.
In a separate process the landing site manager processes incoming landing sites. The given setup before this thesis already had the template in place. Throughout

this work I then replaced the dummy functionality of the landing site manager and the landing node with the final behaviors.
Lastly there the connector threads interact with the flight controller through Mavros and the ROS connector interacting with all the other nodes.

Chapter 5

Methodology

As mentioned above the endeavour of this thesis was to put together a front to back landing procedure utilizing an existing vision based landing site detection pipeline and an autonomous framework. In order to do this, all individual software instances needed to be ready in order to combine them and achieve autonomous landing in unknown terrain.

5.1 Stereo Camera Depth Perception

The autonomous framework[5] allows us to fly independent missions at cruise altitude of 100m+. The structure from motion approach captures 3D information during traversal as its adaptive baseline allows it to perceive high quality depth information also at such high altitudes. This information can be used by LSD in order to detect landing sites during mission.
At low altitudes SFM works as well but surrounded with obstacles, the need for lateral motion poses significant risk. This is because the drone does not retain any hazard information due to the limitations of computational complexity present for mars flights.
In the following a qualitative analysis of the comparison of SFM and stereo camera depth is layed out:

5.1.1 Lateral Motion

As already mentioned above the need for lateral motion in itself is an undesirable necessity for a rotorcraft in unknown terrain.
In this setup the structure from motion approach is based on a keyframe buffer which needs to be filled with image-pose pairs at different horizontal positions in order to start acquiring depth information. The current setting in the implementation Domnik et al. [6] uses 6 keyframes. Therefore for a single point cloud it is necessary to move laterally 6 times in order to start perceiving depth.

5.1.2 Software vs Hardware Depth Perception

Structure from Motion, being a software node that relies on camera poses supplied by a state estimator, is by design subject to inaccuracies. A depth node based on a stereo camera on the other hand works with a fixed rigid baseline between the camera views. Thus for low altitude flights that bear the danger of collision, a more robust hardware approach is preferred.

5.1.3 DEM Conversion

As described in section 4.2.2 the multi-resolution DEM used for depth aggregation in LSD is based on Optimal Mixture of Gaussian cells and thus converges over time. According to section 4.2.2 the landing sites chosen are likely on terrain with low uncertainty. Because of this landing sites are more likely to be detected and have in general a better quality when the terrain perceived has been viewed.
When a landing site has been selected we need to make sure that the landing site is actually correctly detected and of good quality. For this we would like to (re-)detect landing sites on rather converged terrain. Structure from Motion needs constant lateral motion for this. A stereo camera depth node simply hovers in place for any given amount of time.

Efficiency

All in all the stereo camera setup allows us to perceive a landing site at course altitude and after having traversed horizontally to that location, we can simply descent to a stereo camera friendly altitude for the verification. Compared to repeated lateral coverage of the area in question this is a huge increase in efficiency.

5.1.4 Theoretical Analysis

When it comes to depth perception the obvious drawback of a stereo camera is its limited baseline. It only perceives depth accurately for objects within a certain proximity to the lense.
Assuming a perfectly calibrated and rectified camera there is still always an inaccuracy in the depth estimation arising from the disparity error.
From a given disparity estimate the depth error is derived as follows:

$$z = \frac{f \cdot b}{d} \tag{5.1}$$

Where b is the z is the depth estimate, b is the baseline, f is the focal length and d is the disparity value.
Taking the derivative of z w.r.t. d we get

$$\frac{\partial z}{\partial d} = -\frac{f \cdot b}{z^2} \tag{5.2}$$

And substituting (eq. (5.1)) we get:

$$\partial z = \frac{z^2}{f \cdot b} \partial d \tag{5.3}$$

Where the sign was left away as for our application there lies equal danger in a point being perceived too close and too far away.
For the maximum altitude given a maximum allowable depth error this yields:

$$z_{\max} = \sqrt{\frac{\Delta z_{\max} \cdot b \cdot f}{\Delta d}} \tag{5.4}$$

Where Δz is the depth error and Δd the disparity error.
The stereo camera mounted on the drone in JPL's aerial vehicle lab had a baseline of about 10cm and a focal length of 256.
With these properties and estimating a subpixel precision disparity error of 0.25 pixels

5.1.5 Qualitative Practical Analysis

5.2 Landing Site Properties

Before this work the output of the landing site dection algorithm was merely the location of a found landing site. However as described in section 4.2.2 the landing site detection algorithms segments hazards based on roughness and slope. Subsequently it considers the size of a landing site as well as the uncertainty associated with a certain selected location.
Simply outputting the location of a landing site is therefore a waste of information when so many characteristics are at hand to make an informed selection.
I decided on the following properties to be output alongside the site's location:

- Uncertainty
- Roughness
- Size
- Obstacle Altitude

The final landing site detection output is a custom landing site ROS message containing the above mentioned characteristics of the detected spot.

5.2.1 Roughness

The roughness value the exact value already used for the hazard segmentation step in the landing site detection.

5.2.2 Uncertainty

The uncertainty value is also a product of the landing site detection algorithm. It denotes the averaged uncertainty across the area around a given landing site. The uncertainty of a single map cell denotes the stereo depth error estimates merged over time.

5.2.3 Size

To determine the size of a landing site, the landing site detection algorithm performs a distance transform on the created landing site map in order to find the closest non-landing site for any found landing site. This returns the radius of the largest valid landing circle around a landing site. Calculating the physical value, the metric radius is returned as the size of a landing site.

5.2.4 Obstalce Altitude

The obstacle altitude was newly introduced in this work. It defines the currently highest point of the aggregated DEM's highest resolution layer. As no actual object detection is performed and no hazard information is retained in this visual pipeline, this value serves the autonomy as an indication of the obstacles heights to avoid in the vicinity of a certain landing site. More on this in section 5.3.1.

5.3 Autonomous Landing Procedure

With the enhanced structure of the LSD output and having implemented a stereo camera as a low altitude alternative to SFM, the main contribution of this work could be faced.

5.3.1 Landing Site Handling

As shown in section 4.3 the autonomy is structured in a hierarchical and modular way. The current state of the state machine determines the specific task to be executed in that state's execution node. As the landing sites are constantly received alongside the mission tasks performed, they have to be processed in a separate thread. This is handled by a landing site manager (LSM) singleton class which ranks the landing sites according to a loss function and stores them in a heap buffer.

Landing Site Heuristic

As mentioned in section 5.2 the autonomy receives landing sites with the following properties:

- Position
- Roughness
- Uncertainty
- Size
- Obstacle Altitude

From these properties the characteristics that comprise the final heuristic are:

- Current Distance to Drone
- Roughness
- Uncertainty
- Size
- Verification Altitude

Current Distance to Drone
Each iteration the current distance to the drone's position is calculate for each retained landing site. The distance is then normalized by dividing it by the cruise altitude which is 100m. Note that in practice landing sites fell off when farther away than 100m which yields a valid normalization.

Roughness
The roughness property is the unaltered roughness value received from LSD. It is already normalized and enters the loss function as it is.
Uncertainty
The same holds for the uncertainty. It is already normalized by design and enters the loss function unaltered.
Size
Analogous to the roughness and uncertainty properties the size comes from the landing site detection directly. However unlike the two preceeding properties it is not normalized but simply denotes the metric radius of the largest circle fitted around a given landing site. This is achieved in LSD by performing a distance transform on the landing site image.
In order to normalize this value the maximum landing site size is retained and each landing site's size is divided by it in order to achieve the normalized size indication.
Verification Altitude
A site's verificaiton altitude is the smallest vertical distance between the drone and the landing site at which that site was (re-) detected. Similar to the uncertainty metric the verification altitude indicates how certain we can be about a detected landing site as spots detected at lower flight altitudes are more likely correct due the reduced depth error.

Landing Site Input Processing

The landing site manager receives the landing sites from the ROS connector within the autonomy.

Bibliography

[1] S. Withrow-Maser, W. Johnson, L. Young, H. Cummings, A. Chan, T. Tzanetos, J. Balaram, and J. Bapst, "An advanced mars helicopter design," *Icarus*, vol. 319, no. 745-769, 2019.

[2] D. S. Bayard, D. T. Conway, R. Brockers, J. Delaune, L. Matthies, H. F. Grip, G. Merewether, T. Brown, and A. M. San Martin, "Vision-Based Navigation for the NASA Mars Helicopter," 2019.

[3] M. Offermann, J. Delaune, and R. Brockers, "Map-based localization for micro air vehicles autonomous navigation," 2021.

[4] P. Schoppmann, M. Domnik, G. Kubiak, T. Tzanetos, P. F. Proenca, J. Delaune, and R. Brockers, "Autonomous safe landing site detection for a future mars science helicopter," 2021.

[5] L. Di Pierno, R. Brockers, and R. Hewitt, "Autonomous Long-Range Flight Execution for Future Mars Rotorcraft," 2024.

[6] M. Domnik, P. Proenca, J. Delaune, J. Thiem, and R. Brockers, "Dense 3D-Reconstruction from Monocular Image Sequences for Computationally Constrained UAS," 2021.

[7] P. Schoppmann, P. F. Proenca, J. Delaune, M. Pantic, T. Hinzmann, L. Matthies, R. Siegwart, and R. Brockers, "Multi-Resolution Elevation Mapping and Safe Landing Site Detection with Applications to Planetary Rotorcraft," 2021.

[8] P. F. Proenca, J. Delaune, and R. Brockers, "Optimizing Terrain Mapping and Landing Site Detection for Autonomous UAVs," 2022.