COMPARING DIFFERENT FEATURE ALGORITHMS FOR MONOCULAR VISUAL ODOMETRY AND SENSOR FUSION TECHNIQUES IN ROVER NAVIGATION

ZUZANNA FILIPECKA

THESIS SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
BACHELOR OF SCIENCE IN COGNITIVE SCIENCE & ARTIFICIAL INTELLIGENCE

DEPARTMENT OF
COGNITIVE SCIENCE & ARTIFICIAL INTELLIGENCE
SCHOOL OF HUMANITIES AND DIGITAL SCIENCES
TILBURG UNIVERSITY

STUDENT NUMBER

2051164

COMMITTEE

dr. Sharon Ong
dr. Giacomo Spigler

LOCATION

Tilburg University
School of Humanities and Digital Sciences
Department of Cognitive Science &
Artificial Intelligence
Tilburg, The Netherlands

DATE

July 12, 2023

WORD COUNT

7199

ACKNOWLEDGMENTS

I want to deeply thank my supervisor dr. Sharon Ong for her support and guidance throughout this thesis process.

COMPARING DIFFERENT FEATURE ALGORITHMS FOR MONOCULAR VISUAL ODOMETRY AND SENSOR FUSION TECHNIQUES IN ROVER NAVIGATION

ZUZANNA FILIPECKA

Abstract

This thesis addresses the need for accurate self-localization methods in the rapidly growing field of space exploration and robotics. Monocular visual odometry, a technique underpinned by feature detection and description algorithms, is explored as a vital solution for navigation. As the initial step in visual odometry, feature detection and description enable the system to track unique points across multiple frames, thus modeling the robot's path. The robustness of various feature extractors - BRISK, ORB, and A-KAZE - in the visual odometry process is tested. The Visual Odometry results are then compared to sensor fusion results incorporating an Inertial Measurement Unit (IMU) and Wheel Odometry. The results revealed A-KAZE as the superior algorithm based on our evaluation metrics, outperforming BRISK and ORB. It is observed that visual odometry alone does not achieve sufficient performance, however, sensor fusion forms a reliable baseline as a self-localization system. The findings suggest the effectiveness of A-KAZE as a feature extractor and emphasize the importance of sensor fusion in achieving accurate robot localization.

1 DATA SOURCE, ETHICS, CODE, AND TECHNOLOGY STATEMENT

1.1 *Source/Code/Ethics/Technology Statement Example*

Data Source: Both Long Range Trajectories on Mound Etna Dataset and Affine Covariant Features Oxford dataset have been acquired from the DLR German Aerospace Center, and the University of Oxford respectively, through an online request. Parts of the code have been adapted and reused from Work on this thesis did not involve collecting data from human participants or animals. The code Visual Odometry and Image Matching was adapted as a base from open-source GitHub (Uoip, 2016) repository, and Google Colab code (Kennerley, 2021). The sources are given in the code component. The original owner of the data and code used in this thesis retains ownership of the data and code during and after the completion of this thesis. However, the institution was informed about the use of this data for this thesis and potential research publications. To perform spell-checking and grammar, Grammarly tool has been utilized. No other typesetting tools or services were used.

2 INTRODUCTION

2.1 *Project definition*

This research aims to investigate the accuracy of different feature extraction and sensor fusion methods in monocular visual odometry using the Moon Analogue Navigation Dataset on Mount Etna. The term "odometry" concerns the use of data points from motion sensors to estimate the position of the robot through translation and orientation. For this research, the visual odometry technique is used. VO is an estimation process of a robot's pose using a stream of images coming from one or more cameras that are fixed to the autonomous object (Scaramuzza & Fraundorfer, 2011). It is an important technique for navigation, especially in environments where GPS is unavailable. The study aims to determine the most effective feature detection and description method among ORB, BRISK and A-KAZE to provide insights into their accuracy for the planetary rover use case. They will be assessed by evaluating the localization estimates with the DGPS positions given in a dataset. In addition to examining visual odometry, this thesis presents an alternative navigation solution that fuses data from Inertial Measurement Units (IMU) and wheel odometry sensors, thereby approximating robot localization. This fusion-based approach presents an alternate way of localization, fostering a robust comparison with visual odometry techniques. The findings of this study offer valuable insights into the most proficient feature detection methods, and facilitate an illustrative comparison with localization achieved through the fusion of IMU and wheel odometry. This comprehensive evaluation seeks to pave the way for enhanced robotic self-localization using computer vision technologies.

2.2 *Motivation*

During the past few years, there has been a renewed interest in robotic missions to explore the Moon and Mars, highlighting the need for autonomous robots due to the limitations of past practices that have been deemed insufficiently scalable and sustainable for future space exploration (Nesnas, Fesq, & Volpe, 2021). Such robots perform a series of tasks like perception, cognition, motion control, and localization. The main objective for robot navigation is location estimation through external sensors and cameras (Panigrahi & Bisoy, 2022). For that goal, multiple navigation methods have been explored, among which visual and inertial methods were used substantially. As the lunar environment is a GPS-denied area, traditional navigation methods are unreliable, therefore it is important to choose a sufficient system. In such circumstances, alternative sources of self-localization become crucial, and methods such as Visual Odometry stand out as a promising solution, by localizing the robot through visual inputs (Huang, 2019).

Moreover, from the scientific perspective, several advantages of robotic missions on planetary exploration might arise such as opportunities for research in the areas like astronomy, astrobiology, or life sciences (Crawford et al., 2012). Enabling those objectives to be successful, requires an improvement of the rover's localization, as it is a primary step for all other discoveries. For this type of exploration, there is an aim to achieve the best accuracy using the least hardware possible and the least computationally costly software. Therefore, in this paper, Monocular Visual Odometry and sensor fusion method are explored as the promising low-cost and effective solution (Z. Lu, Liu, & Lin, 2022). As the software can be computationally expensive, there are different areas for improvement. One of them is the first step of visual odometry – feature extraction and description. Different methods will be compared in terms of their accuracy.

While it is typical to employ sensor fusion by combining Inertial Measurement Unit (IMU) and wheel odometry data, given its advantages such as providing a metric scale and compensating for visual track losses (Qin, Li, & Shen, 2018), this thesis adopts a distinctive approach. Instead of following the widespread practice of fusing IMU data with wheel odometry and visual odometry, the initial emphasis is on evaluating visual odometry on its own merits. This approach then extends to compare its performance with the more traditional fusion of IMU and wheel odometry. This shift in perspective is driven by the identified need for comprehensive research in feature extraction, matching, and sensor fusion techniques tailored for planetary rover applications. The primary objective is to identify the most precise feature detection and motion estimation methodologies for rover navigation, thereby accelerating advancements in the domain of planetary exploration by leveraging the power of optimized visual odometry.

2.3 *Research questions and findings*

RQ1 *What is the best feature algorithm in terms of robustness for image matching based on the benchmark Oxford dataset and Mount Etna Dataset?*

One of the best ways to measure and compare the performance of feature detectors is through feature matching across image frames. This process is based on effectiveness in finding the corresponding points of interest between images. To compare the results from this study, the Oxford data set is used as ground truth and is compared with the Mount Etna primary dataset. The matching ability is evaluated based on the base image and its variations under different conditions such as the introduction of a random shadow, fog (blur), brightness, and rotation by 180 degrees, respectively, and between the base image and its subsequent frames.

RQ2 *To what extent can Monocular Visual Odometry determine an accurate trajectory estimation in Mount Etna Dataset using different feature extraction methods?*

Visual odometry can help with rover localization in GPS-denied environments. The Monocular VO is more efficient in terms of hardware and costs, thus the idea of the evaluation of this method, even though it might lead to higher errors (Laîné et al., 2016)). Here, different feature extraction methods are proposed and evaluated based on their trajectory estimations.

RQ3 *To what extent can sensor fusion of Inertial Measurement Unit (IMU) and Wheel Odometry (WO) perform in terms of the rover's localization accuracy?"*

The initial phase of the research will yield specific accuracy and error measurements, demonstrating the deviation of the plotted path from the original rover traverse as given in the dataset. As localization accuracy is crucial for planetary rovers operating in extraterrestrial environments, the aim is to achieve as high precision as possible. Therefore, a comparison of different measurements, such as IMU and WO sensor fusion, is also undertaken. This serves as a baseline for subsequent comparison with visual odometry, similar to the role of the DGPS provided in the dataset.

After comparing A-KAZE, ORB, and BRISK feature extractors, it was observed that A-KAZE and ORB exhibited comparable robustness in terms of matching abilities. However, BRISK and ORB showcased similar performance in terms of time efficiency, leading to the conclusion that ORB is the most fitting choice for this component. When selecting a feature detector for Visual Odometry, robust performance is essential, and yet again, A-KAZE proved superior, with the least degree of error among the compared methods. Despite these findings, the research posits that relying solely on Monocular Visual Odometry may not provide an accurate solution for self-localization of autonomous rovers. Therefore, an additional comparative analysis of pose estimation efficiency was conducted on the sensor fusion results, which yielded the best approximation of the trajectory.

3 RELATED WORK

3.1 *Visual odometry for Rovers*

Autonomous mobile robots require precise localization to achieve a proper working navigation system. There are many variants for calculating the position of a vehicle such as – wheel odometry, inertial navigation systems (INS), a global positioning system (GPS), or Visual Odometry (VO) (Aqel, Marhaban, Saripan, & et al., 2016). VO focuses on estimating the object's position by tracking the motion of the unique features between a sequence of images and can be more accurate than most of the localization techniques having a relative position error between 0.1-2 percent (Scaramuzza & Fraundorfer, 2011). Those images can be acquired by using stereo, and monocular cameras, as well as omnidirectional types of them and RGB-D cameras (Aqel et al., 2016). The Stereo vision was used in the widely known NASA's twin Mars Exploration Rovers - Spirit and Opportunity. It has been one of the most extensive real-life applications since almost all VO applications are still tested in simulations on Earth. VO has been described as an efficient tool for cases like vehicle safety, executing challenging drive approaches, and enabling more autonomous capabilities. Even though the approach was insightful, stereo-based visual odometry was not the only component responsible for the long success of the mission (Chen et al., 2019). On top of that utilizing stereo cameras is more expensive, needs more calibration activities, and can lead to errors in the ego-motion estimation process (Kitt et al., 2011). Researchers have also explored the use of monocular omnidirectional cameras specifically for the lunar analog environment, because of the need for low computational and hardware costs (Laîné et al., 2016). In that experiment, feature-based VO was combined with a template-matching algorithm to serve as the input to the SLAM technique. Again, the visual odometry component alone had to be optimized by the SLAM algorithm to achieve a relatively effective localization system. Therefore, related work suggests that results obtained solely from Visual Odometry might not be fully sufficient for a complex environment such as Mars or Moon.

3.2 *Visual features for monocular odometry application*

In Visual Odometry (VO), the feature-based method is described as preprocessing the extracted features from the images to track or match them. Since monocular VO doesn't account for scale estimation and other factors, attention is brought to the steps involving features. (Scaramuzza & Fraundorfer, 2011) provide a comprehensive review of ego-motion estimation methods. They discuss three distinct approaches: 2D-to-2D, which estimates relative pose using 2D features; 3D-to-3D, where the pose is recovered from 3D features; and 3D-to-2D, which involves re-projecting 3D features between cameras. These methods offer different perspectives for estimating ego motion based on the type and availability of features utilized. The performance of feature selection has also a heavy influence on the results of pose estimation. Two main aspects of its performance are mentioned, feature detection and outlier removal (Nguyen & Lee, 2019). Firstly, research on many proposed feature detectors has been conducted, but among the most common are SIFT, and SURF. Although both have proven to be successful in many applications, they also have drawbacks, and their performance can be dependent on the application input. Both SIFT and SURF have a large computational complexity, while ORB on the other hand is thought to have the highest speed (Karami, Prasad, & Shehata, 2017). On top of that, SIFT and SURF are currently patented which leads many researchers to choose other alternatives. Thus, algorithms like ORB or BRISK gained popularity in visual odometry research (He, Zhu, Huang, Ren, & Liu, 2019). For a broad and exhaustive comparison, A-KAZE feature detectors is also tested. From previous experiments it has demonstrated the best trade-off between motion estimation accuracy and computation efficiency, therefore it is also chosen for the comparative framework in this theses (Chien, Chuang, Chen, & Klette, 2016). In terms of the planetary rover application, a new algorithm was proposed that dynamically switches between different feature detectors leading to a great performance in stability and accuracy of feature detection. (Otsu, Otsuki, Ishigami, & Kubota, 2012). Because in the aforementioned process, features are detected through matching, they are subjected to a lot of noise. Outlier rejection is a really important step because wrong matches between frames can lead to big errors. In odometry research, this issue is most commonly tackled by RANSAC algorithm (Kitt et al., 2011). Lastly, feature distribution utilizing the Bucketing technique (Geiger, Ziegler, & Stiller, 2011), the age of the features, meaning how long they are being tracked

(Cvišić & Petrović, 2015) can also lead to reducing noise in feature selection and further improve the efficiency of the feature tracking.

3.3 *Fusion of Inertial Measurement Unit and Wheel Odometry*

Within the literature, several approaches have been identified to fuse sensors and improve navigation results. One such approach is based on the fusion of inertial measurements from an Inertial Measurement Unit (IMU) and Wheel Odometry (WO). Several algorithms are available for sensor fusion, with Extended Kalman Filter being a prevalent base due to its low computational complexity. For instance, the Multi-State Constraint Kalman Filter (MSCKF) has been employed to estimate the localization for mobile robots, offering potential insights for rover applications (Heo, Cha, & Park, 2017). Moreover, the necessity for appropriate feature selection in sensor fusion has been emphasized, as corrupted or missing data can lead to significant errors. As such, a feature selection framework for sensor fusion has been introduced, encompassing two strategies - soft fusion and hard fusion (Chen et al., 2019).

In autonomous mobile robotics, particularly in space exploration, accurate self-localization is crucial. Visual Odometry (VO), a technique that estimates an object's position by tracking unique features across sequential images, has been employed in rovers like NASA's Mars Exploration Rovers - Spirit and Opportunity, showing a potential for accuracy. However, sole reliance on VO may not be sufficient for complex environments such as Mars or the Moon. Feature-based methods in VO, which involve preprocessing extracted features for tracking or matching, are key, with detectors like SIFT, SURF, ORB, BRISK, and A-KAZE showing varying degrees of success. To optimize performance, noise reduction and proper feature selection are critical. To enhance results, sensor fusion combining VO with other sensors, like Inertial Measurement Units (IMUs), has been adopted. Techniques such as the Extended Kalman Filter, used in Visual Inertial Odometry (VIO), offer promising results due to their low computational complexity. Nevertheless, in sensor fusion, accurate feature selection is imperative to prevent significant errors from corrupted or missing data.

4 METHODS

The Methods section provides a concise overview of the experimental procedures. It begins with a comprehensive description of the dataset used, including any necessary adjustments made for experimental suitability. Subsequently, the visual odometry approach is detailed, covering key steps

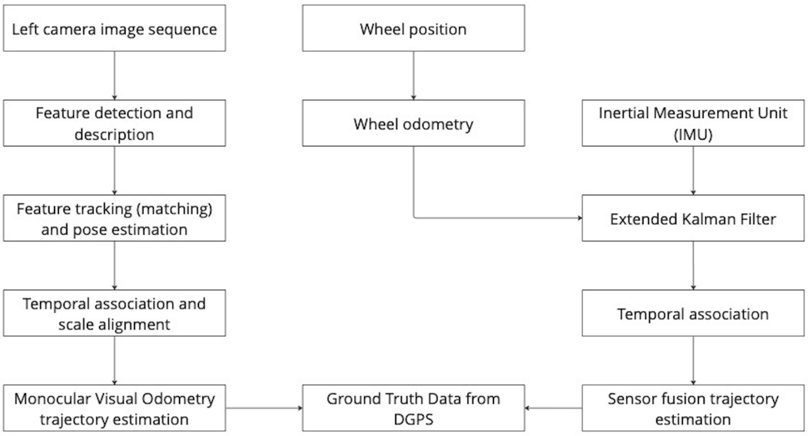
Left camera image sequence
Wheel position
Feature detection and description
Wheel odometry
Inertial Measurement Unit (IMU)
Feature tracking (matching) and pose estimation
Extended Kalman Filter
Temporal association and scale alignment
Temporal association
Monocular Visual Odometry trajectory estimation
Ground Truth Data from DGPS
Sensor fusion trajectory estimation

Figure 1: The figure represents the architecture of the research framework, consisting of three segments. The visual odometry process transforms camera images into trajectory estimates. Wheel odometry readings, combined with IMU data in an Extended Kalman Filter, yield sensor fusion trajectory estimation. Finally, these estimations are compared with DGPS ground truth for analysis.

such as ingesting the left camera sequence, feature detection, tracking, and pose estimation. The section then discusses alignment techniques and post-processing methods employed to refine visual odometry results and enhance accuracy. Furthermore, the integration of sensor fusion techniques for incorporating IMU and Wheel Odometry data is explored, highlighting their contributions to robust estimation, as well as their respective temporal association. The software tools, packages, and evaluation criteria used for performance assessment are also described. The whole framework is presented in (Figure 1).

4.1 *Dataset description*

The dataset used for this study was created by DLR (German Aerospace Center) at a planetary surface analog test site on Mount Etna, Sicily, Italy. It is a publicly available dataset and was released in 2017. The dataset contains the measurements of interest for this paper conducted by the rover sensors and cameras such as images from the stereo camera, sensor measurements from the Inertial Measurement Unit (IMU), Wheel Odometry data points, and differential global positioning system (DGPS) data (Vayugundla et al., 2018). For the aim of this research, the visual inputs only from the left camera are taken to assess the monocular visual odometry. All the data is shared in the ROS bag format consisting of grey-scale images. In addition to that, camera intrinsic and extrinsic parameters were provided for the output calibration aims. The assumptions for Visual Odometry consist of substantial illumination in the environment, the dominance of static scenes over moving objects, enough texture to enable feature tracking and a great scene overlap between the frames. All of those assumptions are fulfilled for this project, thus the choice for this dataset.

There are 53335 images shared which were a visual representation of the trajectory of around 834m. The timestamps which are later used for comparison with ground truth were extracted from the ROS bag as well. Another necessary component for evaluating the results from the dataset is the ground truth file recorded by the differential GPS model combined with a reference DGPS station. There was an interruption during the collection of that data closer to the end of the run, which leads to the discarding of these entries in this paper. Apart from that, the position of the camera in the first 4m changes substantially, which would require a great amount of information regarding camera transformations. Since they are not given, the beginning of the dataset is also discarded. Because of the primary high error in trajectory estimation and the need for multiple alignments and transformations, only the first 10 thousand images and their corresponding ground truth entries were used for the experiment. Both IMU and wheel odometry were reported to have two types of information, angular velocity, and linear acceleration. Wheel odometry used for sensor fusion was given in the text file representing timestamped twist.

Additionally, to this primary dataset for one of the goals of the paper five publicly available images were used as a ground truth for image matching and feature detector performance, as they were thought to have a great degree of variety of image content. For comparison of feature detectors' performance, five photos were also chosen accordingly from the Mount Etna Dataset based on the variety of the terrain recorded by the camera.

4.2 *Visual odometry*

Feature-based Visual odometry was chosen as a primary source of pose estimation for autonomous robot localization. As mentioned before, the need to reduce the costs of the hardware led to the exploration of monocular visual odometry. The primary objective of this part of the project is the estimation of the pose of the vehicle by examining the shifts that motion introduces on the images of its onboard cameras. This can be achieved by connecting the transformations between successive image frames, thus facilitating the estimation of the current robot pose. Achieving a considerable

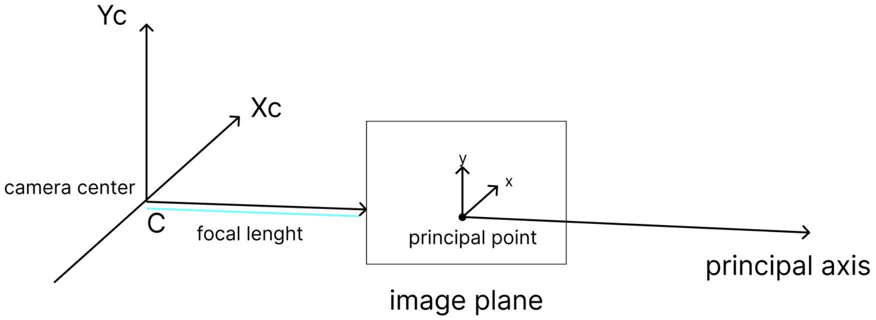

Figure 2: The figure presents the principal components of camera calibration parameters. It includes the focal length, denoted as (fx, fy), that measures the lens to image plane distance, and the optical center (cx, cy), specifying the coordinates where the optical axis intercepts the image plane.

result requires using intrinsic and extrinsic camera parameters. Intrinsic camera parameters describe the internal geometry and optical properties which are needed for visual odometry calculation. Among those, there is a focal length *(fx, fy)*, which represents the distance between the lenses and the image plane, and the optical center *(cx, cy)* represents the coordinates of the principal point, which stands for the intersection of the optical axis with the image plane, as seen in (Figure 2) (Palmieri, Castaldo, & Marino, 2013).

Lastly, among the intrinsic information, there are distortion coefficients *(k1, k2, k3, p1, p2*), which stand for the parameters modeling the radial and tangential distortions caused by the lenses. Apart from intrinsic information, there is also a need for incorporating extrinsic parameters. They, on the other hand, describe the camera's pose with respect to the global coordinate system (Vayugundla et al., 2018). Usually, in the case of stereovision visual odometry applications, they are used to describe the relative position between cameras, but as for monocular applications, they are rather used for converting the relative estimates to the absolute pose for later comparison.

Chosen pipeline consists of a three-stage process. It begins with the detection of unique features within the set of image frames $I_{0:n} = \{I_0, \ldots, I_n\}$ where n is the number of image frames taken at time k, employing robust feature extractors like ORB, BRISK, and A-KAZE. In the subsequent step, the Kanade-Lucas-Tomasi (KLT) Tracker is utilized to monitor the progression of these identified features across consecutive frames. This operation is done by minimizing the pixel intensity disparities within the localized window from one frame to another, tracking 2D feature points in relation to the preceding image (Lucas & Kanade, 1981). The concluding phase involves motion and pose estimation. The full trajectory can be recovered by concatenating all of the image changes. An Essential Matrix (1), capturing the geometric relationship of the frames, is computed (Nguyen & Lee, 2019).

$$T_k = \begin{bmatrix} R_{k,k-1} & t_{k,k-1} \\ \mathbf{0} & 1 \end{bmatrix} \tag{1}$$

For rejecting outliers and improving the reliability of the matrix, the Random Sample Consensus (RANSAC) algorithm is used. Following, the relative camera rotation and translation are recovered by decomposing the essential matrix into possible vectors and choosing the ones that are in front of the camera. Finally, this process outputs the relative X, Y, and Z pose estimation for every pair of images and is written into a TUM (Technical University of Munich) text file with the corresponding timestamp of every image. The format of this output file consists of the following data points: *timestamp of the particular image, tx, ty, tz* being three floating-point numbers representing the translation of the pose in the x, y, and z directions, respectively, and *qx, qy, qz, qw*, being four numbers representing the orientation of the pose as a quaternion, which all have a value of 0 because the orientation was later calculated during the sensor fusion component.

4.2.1 *Feature detection and description*

Feature detection and description are the first steps of the visual odometry pipeline, so they heavily influence the performance of further steps. Feature detection is an activity that focuses on identifying points of interest in the presented scene. Feature descriptors, on the other hand, provide a description of the region around the detected points and, by that, they enable the features for further comparison. Based on that, further motion is estimated between the features from two consecutive frames. This action is also dependent based on the scenery that is being analyzed. In the case of this research, both the environment where the dataset was taken and future sceneries for the planetary rover's application, are thought to be arbitrary in terms of conditions. Since the terrain is not fully explored, it cannot be assumed that it is planar, thus the important role of choosing the best feature detector. Additionally, the changes in the light, scale, and rotation of the environment pose challenges to the feature detector's performance. Under the following conditions of the applications and possible variations of them, scale-space feature detection algorithms are used.

On top of that, scale is an important subject in terms of estimating a pose of an autonomous vehicle from visual input. When executing stereovision odometry, there is a specific baseline based on intrinsic camera parameters used as a reference to recover the scale of the ego-motion (Aqel et al., 2016). In the monocular VO case, there is no such baseline, thus the detectors, which can take scale changes into account, are thought to be a better approach. Two of the most common scale-space feature detectors SIFT and SURF were excluded from this study due to their licensing.

Therefore, firstly ORB (Oriented FAST and Rotated BRIEF) is used, as it is a great alternative to previously mentioned algorithms proposed by (Rublee, Rabaud, Konolige, & Bradski, 2011). This algorithm is based on the key point detector FAST and BRIEF descriptors. After the FAST method identifies the potential key points, ORB uses the Harris corner measure to select the best ones. Later, it calculates a center of gravity for the point area and creates a vector from the point to this center, which is describing its orientation. The main weakness of this approach is the lack of rotational invariance in its descriptor. To overcome this problem, the algorithm computes a rotation matrix using previously computed orientation point (Karami et al., 2017). Secondly, BRISK (Binary Robust Invariant Scalable Keypoints) is utilized. It detects features both on the image plane and on the scale space. To accomplish this, BRISK uses a combination of image scale filters resulting in scale-invariant results. It samples pixel pairs around the point of interest. Long-distance pairs help with finding the orientation and rotation of the point. Short-distance pairs are compared later creating a binary description based on the brightness of pixels (Leutenegger, Chli, & Siegwart, 2011). Lastly, the Accelerated-KAZE (AKAZE) is chosen because of its scale-invariance (Chien et al., 2016). It uses Fast Explicit Diffusion to create non-linear spaces. Feature detection is done by the Hessian Matrix detector. Additionally, the feature description part is executed by the Modified Local Difference Binary algorithm. The choice for these three algorithms underlies the objectives of this paper. They are thought to be reliable for variations in scale, rotation, and condition changes, providing a robust representation of an image.

4.2.2 *Feature matching, tracking, and pose estimation*

After the points of interest are detected from the image, the feature motion analysis can be conducted using two commonly used approaches – feature tracking and feature matching (Scaramuzza & Fraundorfer, 2011). The goal of the first approach is to find some correspondences between features in different images. Therefore, it is used to answer the first research question regarding how efficient and robust different algorithms are for finding those corresponding interest points. Both for ground truth and images from the Mount Etna dataset, key points are extracted and described using ORB, BRISK, and A-KAZE and later they are matched by the Brute-Force (BF) Method in OpenCV. It simply compares the descriptor of one feature in the first part with all the other descriptors from the image that it is being compared with and chooses the best one with the smallest distance (Hamming distance) (Dong, Fan, Ma, & Ji, 2021). After that step, various evaluation steps are performed, specifically the matching abilities which are calculated by dividing the amount of the sorted matched points in the previous frame by the number of features in the current frame. Feature tracking, on the other hand, comes into play when working with continuous visual data. In this project, as visual odometry is based on the sequence of images, this method is preferred. The commonly used Lukas-Kanade method is chosen for tracking the key points. By assuming that the pixel intensities of features remain constant over time, the KLT Tracker calculates the optical flow, and from that the rover pose is estimated.

4.3 *Temporal association and scale alignment for Monocular VO*

Lastly, in order to enable the valid comparison of the Visual Odometry estimation with the reference trajectory the post-processing steps are applied. There was a need for temporal and scale alignment (Vayugundla et al., 2018). As previously mentioned, while using Monocular Visual Odometry, it is not possible to estimate the scale and the distance between the objects in the observed scene. Since the camera provides 2D image measurements, it lacks the possibility to obtain depth information, therefore the scale of the estimated motion is ambiguous (Aqel et al., 2016). Another obstacle that had to be taken care of was the different frequencies at which the DGPS (10Hz) and Visual Odometry (14Hz) were given. Thus to overcome this problem, performing a temporal association for the entries that don't differ by more than 0.02(s), was necessary.

4.4 *Sensor fusion of IMU and wheel odometry*

For comparison with the VO results, the wheel odometry is combined with an Inertial Measurement Unit system. The readings of angular rates and linear accelerations were taken together and fused within the Extended Kalman Filter (EKF) utilizing the robot_localization package. The EKF works based on updating an initial estimate of the entity's state based on the available sensors. This process is based on two main steps. Firstly, the prediction step is used to project the state estimate and error covariance that represents the uncertainty in the prediction. Consequently, the update step computes the non-linear functions and flattens them to obtain an estimation. Those two steps are continuously interleaved leading to refinement of the estimation. In this application, the output of the EKF sensor fusion outputs a 6DOF trajectory estimation. Linear accelerations and angular velocities are able to produce information about the robot's translation motion with X, Y, and Z estimates and rotational motion (roll, pitch, and yaw). The rotational yaw can be estimated by tracking the wheel rotation rates. In the end, for the sensor fusion component, in the same manner, as for Visual Odometry, the temporal association was performed, since the differences in the frequencies of the provided entries were substantial. Lastly, to make sure that the sensor entries are fused in the same reference frame, wheel odometry was transformed from its frame to the IMU frame since it was used as the base link in the provided dataset.

4.5 *Evaluation criteria*

The primary objective of this project is to point out the most efficient detection and sensor fusion methods utilizing monocular visual odometry for autonomous rover navigation. The feature algorithms can be used for several goals such as feature detection, description, matching, and tracking.

The first research question is posed to explore their efficiency in terms of the ability to find key points and match them against two image frames. Thus, for image matching abilities evaluation of ORB, A-KAZE, and BRISK a ground truth Oxford dataset is used. Five primary images from the Oxford dataset (Bikes, Boat, Graffiti, Cars, and Ubc) sequences are taken, as well as five manually chosen photos from the primary dataset representing various conditions. As for the rover navigation application, the trade-off between efficiency and computational and time cost is the most important, it is a leading factor for evaluating the best algorithm. To achieve a reliable and robust evaluation process, the image sequences are subjected to analysis under different evaluation criteria such as processing time, computational cost, and matching capability measured in percentage. All mentioned criteria are measured under different conditions, the primarily given condition, rotation (by 180 degrees), brightness, blur, and random shadow. The evaluation of the algorithms is based on their robustness to different changes in conditions. However, their performance is also dependent on the parameters that they are being subjected to. Each corresponding parameter is shown in (Table 3). To reliably compare the methods, the feature detection and description are done on different values for given parameters.

However, in real-life applications, like visual odometry, feature tracking is also crucial to be performed effectively. Therefore, ORB, A-KAZE, and BRISK are also evaluated on the run over the sequence of images that monocular visual odometry is calculated based on as well, being around ten thousand images. To answer the second research question, the relative pose estimation must be compared with the absolute pose from the DGPS

(Differential global positioning system), which is already provided in the dataset in the text file format. Before the evaluation can be done, there is a need for temporal alignment and scale alignment, which can be executed directly using Evaluation Visual Odometry (EVO) package. For the temporal alignment, the reference ground truth trajectory is aligned with the estimated trajectory by the maximum difference of 0.02(s) in their provided timestamps. Further, scale alignment is performed using (Umeyama, 1991) with scale correction. The accuracy is evaluated based on the Absolute Pose Error (APE) which directly quantifies the deviation between two poses: a reference pose $P_{\text{ref},i}$ and an estimated pose $P_{\text{est},i}$ at timestamp i. The APE at timestamp i, denoted as E_i, is calculated as the relative pose between the estimated pose and the reference pose as shown in notation (2).

$$E_i = P_{\text{est},i} \ominus P_{\text{ref},i} = P_{\text{ref},i}^{-1} \cdot P_{\text{est},i} \in \mathrm{SE}(3) \tag{2}$$

Here, $\ominus$ stands for the inverse compositional operator, which computes the relative pose by taking the inverse of the reference pose and then composing it with the estimated pose. This relative pose provides a measure of the absolute error between the estimated and reference poses at a specific timestamp i (F. Lu & Milios, 1997).

4.6 *Software and packages*

This project was executed on MacOS Big Sur 11.7.4. The main programming language is Python (version 3.9). Visual Studio Code was used as the main interpreter. Additionally, because of the more compatible framework for the sensor fusion component, VirtualBox 7.0.8 was used to create a virtual machine with Ubuntu 18.04 system and Robot Operating System 1 (Noetic version). For that component catkin workspace was created to handle the sensor fusion. A manually created source package was used together with the open-sourced robot_localization package (Moore & Stouch, 2014) which is designed specifically for sensor fusion. All of those elements were put in the catkin workspace where sensor fusion was recorded. Inside the manually created source package, there are respective subsection packages: etna_interface used to manage launch files, configurations, and visual tools such as Rviz (allows for the real-time visualization of the components of the robotic system) and Rqt (graphical interface framework which allows for implementing visual tools). Following, there is also an etna_translator package which takes the IMU and Wheel Odometry data as published by the Rosbag and formats them to topics that are further ingested with the sensor fusion engine. The main packages that were utilized extensively were: OpenCV, NumPy, seaborn, matplotlib, SciPy and scikit-learn.

For both Visual Odometry and sensor fusion evaluation an open-source package was chosen called EVO (Evaluation Visual Odometry).

5 RESULTS

In this section results for three previously chosen feature detector-descriptor algorithms are reported. Firstly, their performance in terms of feature matching is described. Following, trajectory estimation results are presented for different parameters and algorithms, compared between each other and against the sensor fusion component estimation.

5.1 *Performance comparison for feature matching and tracking*

All of the feature algorithms have several advantages and disadvantages under different conditions. Therefore, there is a need for a thorough evaluation of them. Firstly, exploring feature-matching abilities was evaluated under the computational time criterion. In order for those results to be comparable, all the detector's parameters were set such that they detected approximately 300 features. The results are presented as an average time in seconds over 5 different, manually chosen images. As shown in (Table 1) ORB and BRISK, in this case, performed similarly achieving the best results of 0.0144 (s) for ground truth and 0.0196 (s) for the Mount Etna data set, respectively. A-KAZE for both datasets performed worse than the other methods.

Algorithm	Ground truth	Mount Etna
ORB	**0.0144**	0.0247
BRISK	0.0161	**0.0196**
A-KAZE	0.0764	0.1253

Table 1: Average computational time (s) of 5 images for different algorithms for 300 features

A second evaluation criterion that reliably indicates the precision of the algorithms is the *average percentage of matched points* across the frames, also referred to as *repeatability*. In the feature detection manner, this term signifies the consistency of a feature detector in identifying the same features under varying conditions. These conditions can span distinct viewpoints, scales, lighting conditions, and so forth. Specifically in this experiment, features were matched between the initial image and images subjected to four different environmental manipulations, namely random shadow, random fog (signifying blur), brightness alteration, and rotation.

These conditions were selected based on their potential occurrence in a planetary environment.

The results for the repeatability in Ground Truth and Mount Etna are displayed in (Table 2) and (Table 3), respectively. Regarding the shadow condition, a similar performance was noted across all detectors, with an achievement range of 82.4% to 82.6% for ground truth and a range between 73.5-79.4% for Mount Etna. Upon introduction of random shadow across the image, algorithms adeptly detected features, largely ignoring the shadowed area, and subsequently yielded favorable results. The fog condition, simulating possible image blur, presented a greater challenge. A-KAZE emerged as the top performer under these circumstances, recording a match rate of 38.7% in the ground truth, but ORB turned out to perform for Mount Etna yielding 54.8%. However, blur, in general, had a significant impact, leading to lower overall results across all algorithms, with BRISK yielding only an 8.3% match rate. The brightness alteration condition, conversely, did not substantially affect the matching performance. All algorithms consistently produced results exceeding 83% with BRISK being the worst option among the algorithms. Finally, in response to a 180-degree rotation condition, ORB demonstrated a remarkable 100% match accuracy, distinguishing itself under these specific circumstances. In the end, both A-KAZE and ORB were significantly similar in overall performance. In the shadow and brightness condition, even though A-KAZE didn't achieve the best performance, it is chosen over BRISK which didn't detect the features in the areas of the shadow at all.

Table 2: Algorithm Comparison for the Matched Keypoints Percentage on the Ground truth

Algorithm	Conditions			
	Shadow	Fog	Brightness	Rotation
ORB	82.4%	17.7%	92.3%	**100**%
A-KAZE	**82.6**%	**38.7**%	90.6%	97.5%
BRISK	82.5%	8.3%	**92.6**%	76.6%

(Figure 7) describes the figures representing the best matching percentages for each condition for the ground truth Oxford dataset, and (Figure 12) describes the same condition figures but for the Mount Etna Dataset.

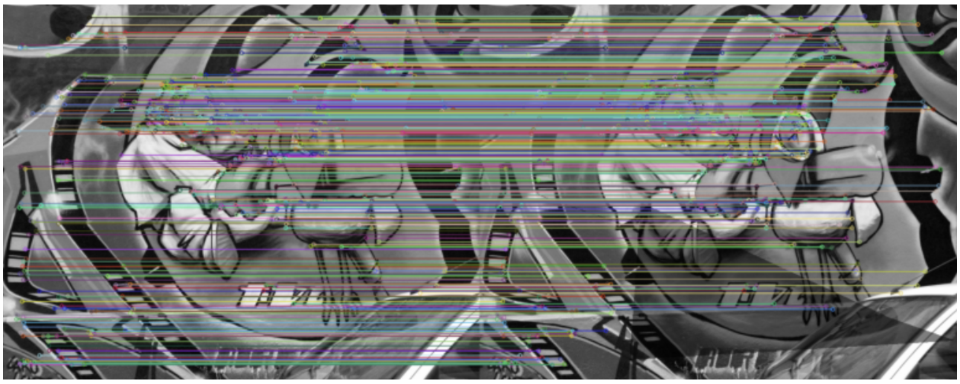

Figure 3: Shadow Condition

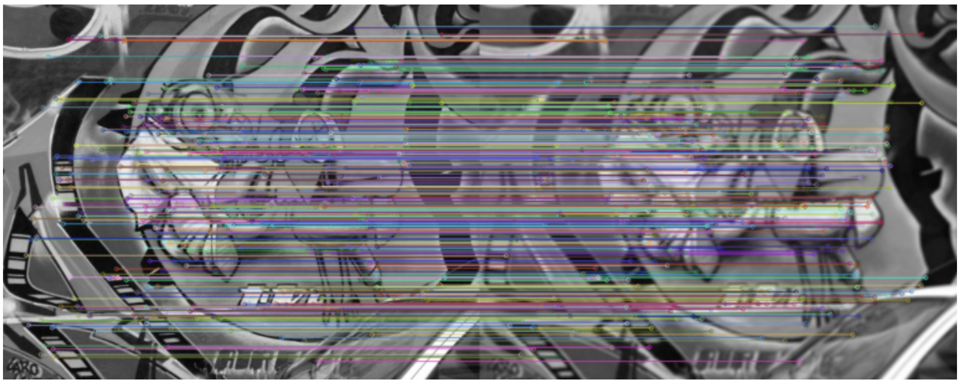

Figure 4: Fog Condition

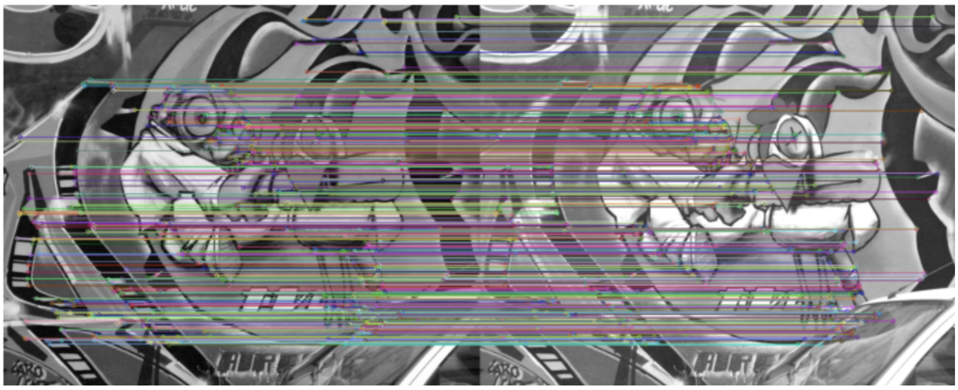

Figure 5: Brightness Condition

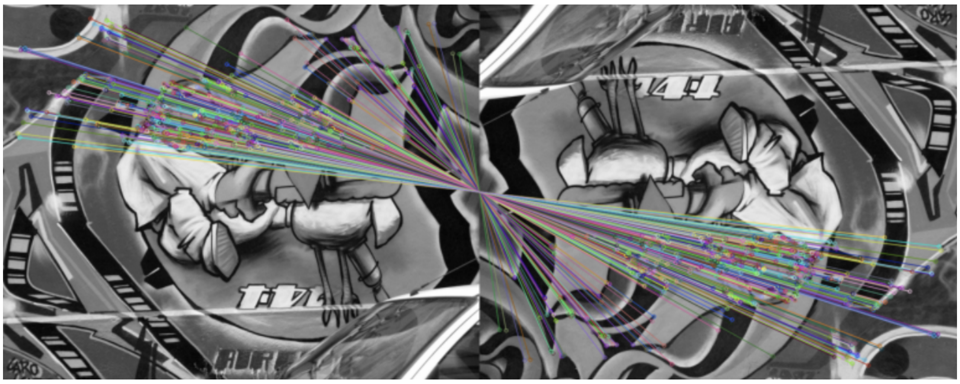

Figure 6: Rotation Condition

Figure 7: Best Matching Percentages Visualizations for Each Condition for Oxford Ground Truth

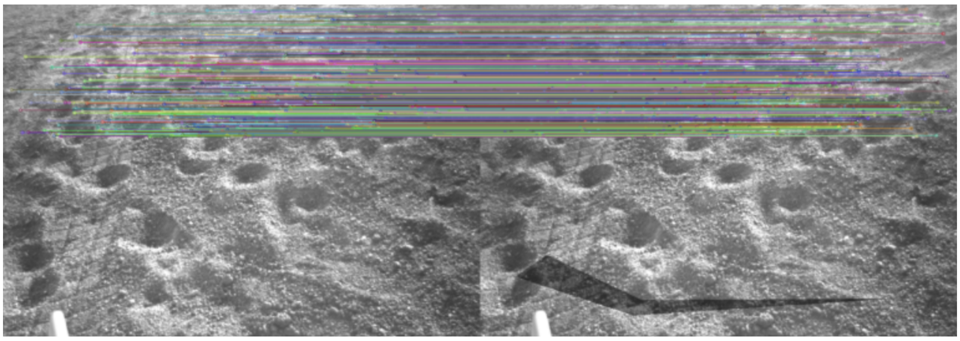

Figure 8: Shadow Condition

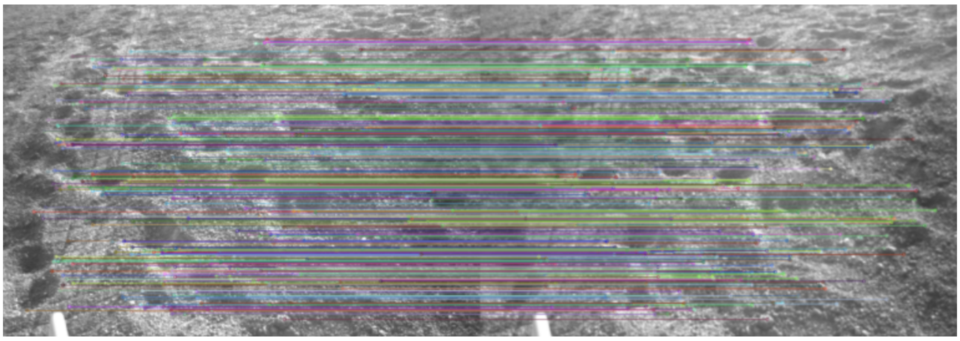

Figure 9: Fog Condition

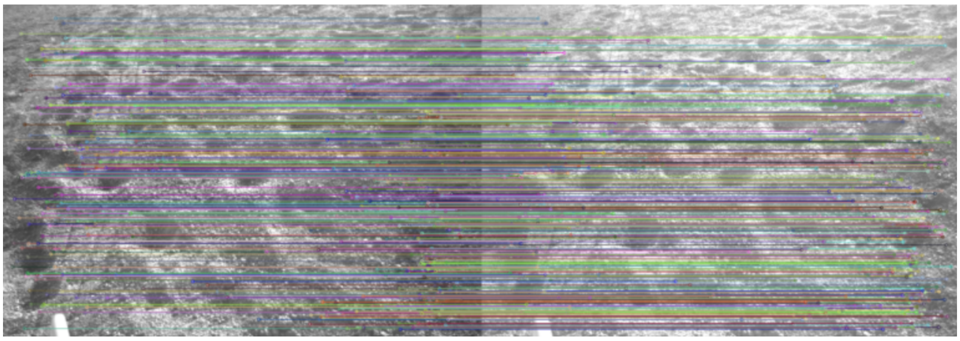

Figure 10: Brightness Condition

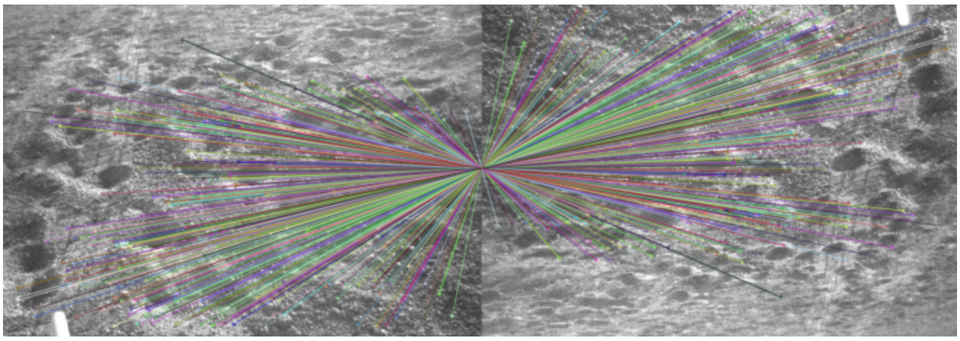

Figure 11: Rotation Condition

Figure 12: Best Matching Percentages Visualizations for Each Condition for Mount Etna Dataset

Table 3: Algorithm Comparison for the Matched Keypoints Percentage on the Mount Etna dataset

Algorithm	Conditions			
	Shadow	Fog	Brightness	Rotation
ORB	73.5%	**54.8**%	84%	**100**%
A-KAZE	**78.4**%	46%	**88.4**%	98.8%
BRISK	79.4%	12.6%	90.8%	82.9%

A supplementary examination for enhancing feature matching efficiency was conducted on ensuing frames. Another objective for monocular visual odometry, a crucial technique for determining camera movement, relies on the precise tracking and alignment of features across distinct images. Consequently, an important aspect of this evaluation process involves an analysis of the tracking of corresponding features between successive features. The initial image was compared with a sequence of 29 subsequent images to assess the consistency of feature representation across these frames. Tests were executed across four separate parameters for each algorithm, each involving approximately 50, 250, 450, and 650 features. Within this comparative framework demonstrated in (Table 4), BRISK yielded superior performance, achieving an unmatched success rate with 50 features. A-KAZE emerged as the runner-up, securing a solid 65.6% match accuracy. ORB, however, yielded a matching percentage roughly half that of BRISK. A noticeable trend was found across the algorithms where the highest matching percentage appeared for the lowest amount of features of 50.

A noteworthy observation pertains to the consistent, but weak negative correlation across all algorithms between the number of features and matching precision. As the feature count escalates, there is a decrease in the percentages of accurately matched points of interest. This trend underscores the potential trade-off between an increased feature representation and the maintenance of matching accuracy, suggesting an additional exploration in the realm of monocular visual odometry. The results are given in (Table 5). For all of the algorithms, there is a weak negative correlation shown by a Pearson correlation coefficient. Although these results are

Algorithm	Parameter	Matching %
A-KAZE	thresh = 0.0073	**65.6**
	thresh = 0.0044	51.3
	thresh = 0.00335	47.9
	thresh = 0.00238	42.4
ORB	nfeatures = 50	**49.7**
	nfeatures = 250	45.4
	nfeatures = 450	44.3
	nfeatures = 650	43.7
BRISK	threshold = 119	**100**
	threshold = 98	79.6
	threshold = 87	61.2
	threshold = 76	55

Table 4: Matching percentage for AKAZE, ORB, and BRISK with different parameters for Mount Etna datset

persistent for multiple image sequences among the dataset, given p-values there is no strong evidence to reject the null hypothesis.

Algorithm	Pearson Correlation Coefficient	P-value
A-KAZE	-0.27	0.1637
ORB	-0.33	0.2
BRISK	-0.24	0.08

Table 5: Pearson correlation coefficients and p-values for AKAZE, ORB, and BRISK on the Mount Etna dataset

5.2 *Trajectory estimation for Monocular Visual Odometry and Sensor Fusion*

Feature-based approach for Monocular Visual Odometry alone turned out to be not effective by itself. The varied structure of the terrain introduced a lot of noise in terms of feature detection. The only distinctive objects that were easily tracked were bigger rocks and sometimes holes in the terrain, but those also led to substantial noise. In the beginning, the visual odometry pipeline was executed on the raw images taken from the dataset. Even though they were provided as already rectified, they had a visible black frame around them. That aspect, introduced a big error for trajectory estimations because the algorithms were treating the edges of the frame as corners, and therefore, the image had to be cropped and visual odometry estimates were provided again.

Sensor fusion of particularly two sensors, IMU and WO resulted in approximately better estimation of the trajectory than solely Visual Odometry. One aspect that has to be taken into account in regard to the noise difference is that this estimation has utilized the Extended Kalman filter, which is responsible for almost invisible noise in the estimation, by taking into account uncertainties and noise present in the system and providing an optimal solution. For this case, scale alignment was not performed, but rather only the temporal association.

Results provided in (Table 6) show different measures calculated based on APE (Absolute Pose Estimation) both for Visual Odometry for each detector and the sensor fusion of Inertial Measurement Unit (IMU) and Wheel Odometry (WO). These results provide the extent of the error of the already aligned trajectories. The output of the spatial alignment was given in the format of rotation and translation alignment matrices calculated from Umeyama's method (Umeyama, 1991). The scale correction was given from this method for each algorithm A-KAZE (0.06), ORB (0.69), and BRISK (0.54), respectively. Already from this, it can be deduced that the trajectory estimation from the A-KAZE algorithm needed the least correction. That also led to the same output for all Max, Min, Mean, Std, and RMSE data points, where A-KAZE had the best results among the feature detectors. The overall accuracy based on Root Mean Squared Error (RMSE) is a commonly used measure of the average error between the estimated and ground truth poses. Their difference for ORB and A-KAZE wasn't substantial, but in the autonomous planetary rovers application and their localization, the accuracy of the estimation is really crucial, which discredits ORB in this comparison. BRISK performed the worst for Monocular Visual Odometry. However, sensor fusion estimates based on IMU and WO entries was found to perform the best out of all the proposed trajectories. Because of the smoothed noise in the sensor fusion component, the standard deviation is the smallest. The RMSE was reported as the lowest with the result of 10.

Table 6: Performance Metrics for Feature Detection Algorithms and Sensor Fusion

Algorithm	**Max**	**Min**	**Mean**	**Std**	**RMSE**
A-KAZE	29.1	1.4	12.5	6.5	14.1
ORB	46.5	1.7	17.7	9.2	19.9
BRISK	70.6	4	34.7	12.3	36.8
IMU + WO	27.5	1.3	8.4	5.5	10

The visual representation of the 2D plots is given in the (Figures 13, 14, 15, 16) where the ground truth is marked as a reference, and the color trajectory is the estimated align trajectory. Additionally, the colors represent the range of error, from minimum, maximum, and mean errors.

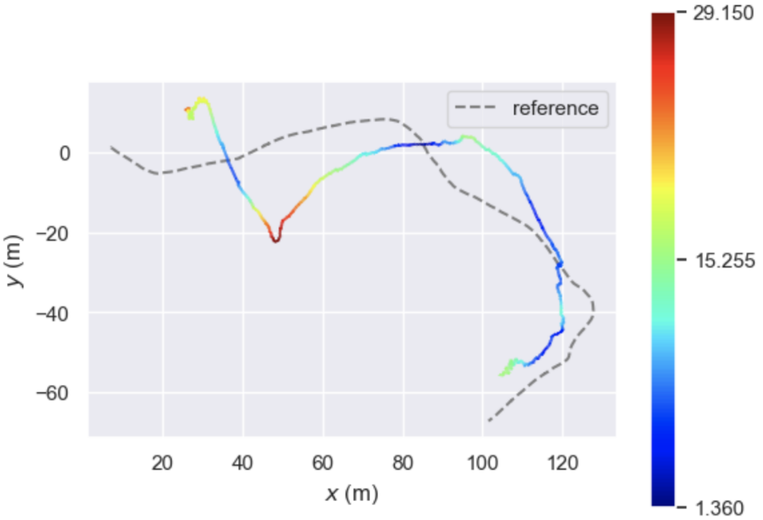

Figure 13: 2D Plot for A-KAZE against a reference ground truth

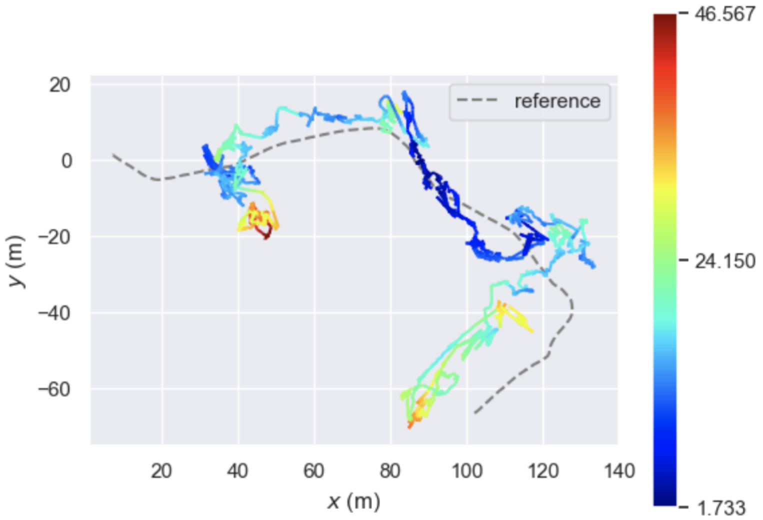

Figure 14: 2D Plot for ORB against a reference ground truth

6 DISCUSSION

The primary goal of this project was to find the best feature extractors techniques for Monocular Visual Odometry and sensor fusion techniques to accurately self-localize the autonomous entity in the planetary application.

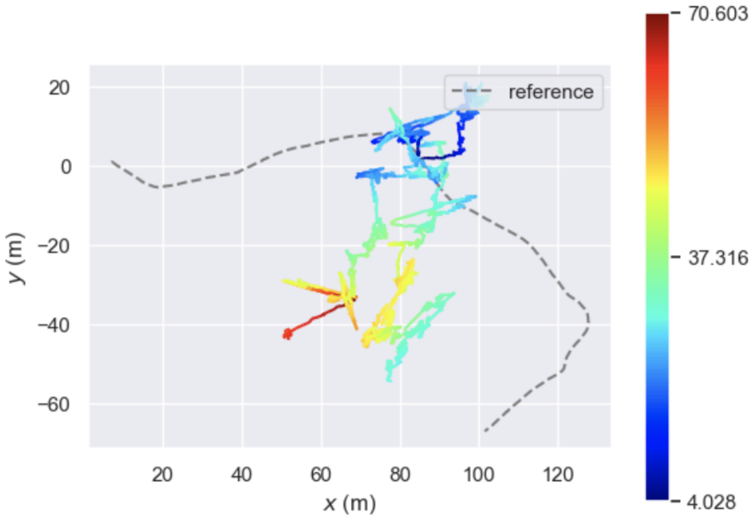

Figure 15: 2D Plot for BRISK against a reference ground truth

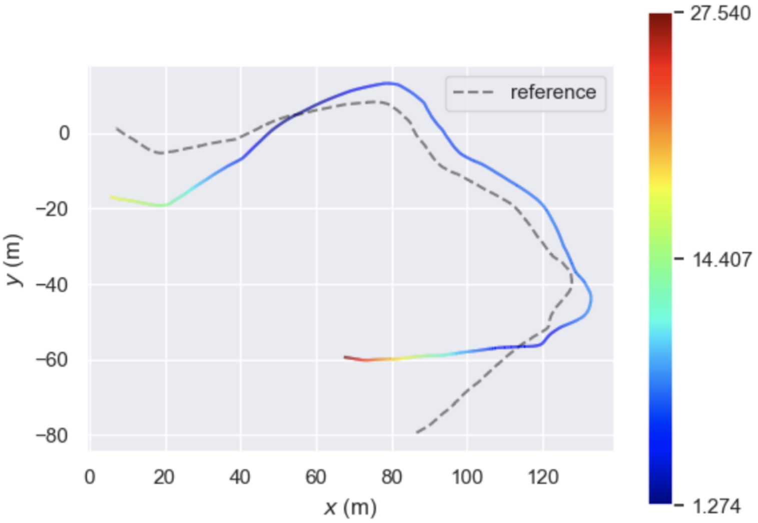

Figure 16: 2D Plot for IMU + WO against a reference ground truth

In this project, various methods were investigated to optimize the accuracy of autonomous vehicle pose estimation. To this end, a feature-based visual odometry approach was implemented. The performance of three distinct feature detectors A-KAZE, ORB, and BRISK was examined under a multitude of conditions and evaluation criteria together with the sensor fusion component as well. Findings from this study are discussed further in the following order of the research questions.

The aim of the first research question was to point out the best feature algorithm in feature matching ability between two different image frames. Even though A-KAZE didn't perform the best under the time efficiency criterion, it has sustained the best performance under different evaluation components, because of its robustness. Therefore, the goal of future work could focus on evaluating the comparison of the A-KAZE algorithm against other state-of-art methods, since it hasn't been explored as widely as ORB, SIFT, or SURF (Karami et al., 2017). Further, the second research question delves deeper into exploring the best feature algorithm for Monocular Visual Odometry application and trajectory estimation. Here, the images captured encompassed a rocky and variable environment that led to non-easily trackable, distinctive objects. Therefore, the algorithms were detecting a lot of unnecessary features and provided a high error in motion and pose estimation. Although this framework posed a lot of obstacles, A-KAZE again turned out to be the most accurate with its trajectory estimation having the least noise among the proposed algorithms. Hence, the exploration for better-suited outlier rejection techniques and feature-tracking retention approaches within this context remain open for future work. Overall, Monocular Visual Odometry, even though it might lead to smaller costs, alone is not the best method for trajectory estimation in planetary applications, because of the inability to estimate scale and track important features in the small variation of the type of terrain. The last research question of this study explores whether commonly used sensor fusion techniques can accurately provide a pose estimation. Combining IMU and WO measures together and fusing them into the Extended Kalman Filter led to the best results for the rover localization, with the lowest degree of error. As other research has shown (Qin et al., 2018), (Heo et al., 2017), and (Bloesch, Burri, Omari, Hutter, & Siegwart, 2017) a combination of Visual and Inertial Systems might lead to even better trajectory approximations, this study also suggests a further exploration of Visual-Intertial Navigation systems for planetary entities.

7 CONCLUSION

In conclusion, feature-based Monocular Visual Odometry performed on gray-scale image frame sequences does not yield the expected results. The grayscale space of the image may be insufficient for this application due to possible illumination or terrain deformations (He et al., 2019). However, the sensor fusion component led to a more reliable approach. Therefore, it is concluded that the combination of Visual and Inertial components may provide the best accuracy. This study contributes to the space and computer vision research field by comparing and pointing out the best feature detector for Monocular VO. It also provides a baseline for further investigation of the most reliable and accurate self-localization and navigation systems.

REFERENCES

Aqel, M., Marhaban, M. H., Saripan, M. I., & et al. (2016). Review of visual odometry: types, approaches, challenges, and applications. *SpringerPlus*, *5*(1), 1897. Retrieved from https://doi.org/10.1186/s40064-016-3573-7 doi: 10.1186/s40064-016-3573-7

Bloesch, M., Burri, M., Omari, S., Hutter, M., & Siegwart, R. (2017). Iterated extended kalman filter based visual-inertial odometry using direct photometric feedback. *The International Journal of Robotics Research*, *36*(10), 1053–1072. doi: 10.1177/0278364917728574

Chen, C., Rosa, S., Miao, Y., Lu, C., Wu, W., Markham, A., & Trigoni, N. (2019). Selective sensor fusion for neural visual-inertial odometry. In *2019 ieee/cvf conference on computer vision and pattern recognition (cvpr).* Retrieved from https://doi.org/10.1109/cvpr.2019.01079 doi: 10.1109/cvpr.2019.01079

Chien, H.-J., Chuang, C.-C., Chen, C.-Y., & Klette, R. (2016). When to use what feature? sift, surf, orb, or a-kaze features for monocular visual odometry. In *2016 international conference on image and vision computing new zealand (ivcnz).* doi: 10.1109/ivcnz.2016.7804434

Crawford, I. A., Anand, M., Cockell, C. S., Falcke, H., Green, D. A., Jaumann, R., & Wieczorek, M. A. (2012). Back to the moon: The scientific rationale for resuming lunar surface exploration. *Planetary and Space Science*, *74*(1), 3–14. Retrieved from https://doi.org/10.1016/j.pss.2012.06.002 doi: 10.1016/j.pss.2012.06.002

Cvišić, I., & Petrović, I. (2015). Stereo odometry based on careful feature selection and tracking. In *2015 european conference on mobile robots (ecmr)* (p. 1-6). doi: 10.1109/ECMR.2015.7324219

Dong, Y., Fan, D., Ma, Q., & Ji, S. (2021). Superpixel-based local features for image matching. *IEEE Access*, *9*, 15467-15484. doi: 10.1109/ACCESS.2021.3052502

Geiger, A., Ziegler, J., & Stiller, C. (2011). Stereoscan: Dense 3d reconstruction in real-time. In *2011 ieee intelligent vehicles symposium (iv)* (p. 963-968). doi: 10.1109/IVS.2011.5940405

He, M., Zhu, C., Huang, Q., Ren, B., & Liu, J. (2019). A review of monocular visual odometry. *The Visual Computer*, *36*(5), 1053–1065. Retrieved from https://doi.org/10.1007/s00371-019-01714-6 doi: 10.1007/s00371-019-01714-6

Heo, S., Cha, J., & Park, C.-G. (2017). Monocular visual inertial navigation for mobile robots using uncertainty based triangulation. *IFAC-PapersOnLine*, *50*(1), 2217–2222. Retrieved from https://doi.org/10.1016/j.ifacol.2017.08.928 doi: 10.1016/j.ifacol.2017.08.928

Huang, G. (2019). Visual-inertial navigation: A concise review. In *2019 international conference on robotics and automation (icra).* Retrieved from https://doi.org/10.1109/icra.2019.8793604 doi: 10.1109/icra.2019.8793604

Karami, E., Prasad, S., & Shehata, M. S. (2017). Image matching using sift, surf, brief and orb: Performance comparison for distorted images. *arXiv*. Retrieved from https://arxiv.org/abs/1710.02726 doi: 10.48550/arXiv.1710.02726

Kennerley, M. (2021, May 21). *A comparison of sift, surf and orb on opencv.* https://mikhail-kennerley.medium.com/a-comparison-of-sift-surf-and-orb-on-opencv-59119b9ec3d0. (Medium)

Kitt, B., Rehder, J., Chambers, A., Schönbein, M., Lategahn, H., & Singh, S. (2011). Monocular visual odometry using a planar road model to solve scale ambiguity. In *European conference on mobile robots.*

Laîné, M., Cruciani, S., Palazzolo, E., Britton, N. J., Cavarelli, X., & Yoshida, K. (2016). Navigation system for a small size lunar exploration rover with a monocular omnidirectional camera. In *Spie proceedings* (Vol. 10011). Retrieved from https://doi.org/10.1117/12.2242871 doi: 10.1117/12.2242871

Leutenegger, S., Chli, M., & Siegwart, R. Y. (2011). Brisk: Binary robust invariant scalable keypoints. In *2011 international conference on computer vision.* Retrieved from https://doi.org/10.1109/iccv.2011.6126542 doi: 10.1109/iccv.2011.6126542

Lu, F., & Milios, E. E. (1997). Globally consistent range scan alignment for environment mapping. *Autonomous Robots*, *4*, 333-349.

Lu, Z., Liu, F., & Lin, X. (2022). Vision-based localization methods under gps-denied conditions. *arXiv*. Retrieved from https://arxiv.org/abs/2211.11988 doi: 10.48550/arxiv.2211.11988

Lucas, B. D., & Kanade, T. (1981). An iterative image registration technique with an application to stereo vision. In *International joint conference on artificial intelligence.*

Moore, T., & Stouch, D. (2014, July). A generalized extended kalman filter implementation for the robot operating system. In *Proceedings of the 13th international conference on intelligent autonomous systems (ias-13).* Springer.

Nesnas, I. A. D., Fesq, L. M., & Volpe, R. A. (2021). Autonomy for space robots: Past, present, and future. *Current Robotics Reports*, *2*(3), 251–263. Retrieved from https://doi.org/10.1007/s43154-021-00057-2 doi: 10.1007/s43154-021-00057-2

Nguyen, H. H., & Lee, S. (2019). Orthogonality index based optimal feature selection for visual odometry. *IEEE Access*, *7*, 62284-62299. doi: 10.1109/ACCESS.2019.2916190

Otsu, K., Otsuki, M., Ishigami, G., & Kubota, T. (2012). Advanced visual odometry for planetary exploration rover..

Palmieri, F. A., Castaldo, F., & Marino, G. (2013). Harbour surveillance with cameras calibrated with ais data. In *2013 ieee aerospace conference.* doi: 10.1109/aero.2013.6496907

Panigrahi, P. K., & Bisoy, S. K. (2022). Localization strategies for autonomous mobile robots: A review. *Journal of King Saud University - Computer and Information Sciences*, *34*(8), 6019–6039. Retrieved from https://doi.org/10.1016/j.jksuci.2021.02.015 doi: 10.1016/j.jksuci.2021.02.015

Qin, T., Li, P., & Shen, S. (2018). Vins-mono: A robust and versatile monocular visual-inertial state estimator. *IEEE Transactions on Robotics*, *34*(4), 1004–1020. Retrieved from https://doi.org/10.1109/tro.2018.2853729 doi: 10.1109/tro.2018.2853729

Rublee, E., Rabaud, V., Konolige, K., & Bradski, G. (2011). Orb: An efficient alternative to sift or surf. In *2011 international conference on computer vision.* Retrieved from https://doi.org/10.1109/iccv.2011.6126544 doi: 10.1109/iccv.2011.6126544

Scaramuzza, D., & Fraundorfer, F. (2011). Visual odometry [tutorial]. *IEEE Robotics & Automation Magazine*, *18*(4), 80–92. Retrieved from https://doi.org/10.1109/mra.2011.943233 doi: 10.1109/mra.2011.943233

Umeyama, S. (1991). Least-squares estimation of transformation parameters between two point patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *13*(4), 376-380. doi: 10.1109/34.88573

Uoip. (2016). *Uoip/monovo-python: A simple monocular visual odometry project in Python.* Retrieved from https://github.com/uoip/monoVO-python

Vayugundla, M., Steidle, F., Smisek, M., Schuster, M. J., Bussmann, K., & Wedler, A. (2018). Datasets of long range navigation experiments in a moon analogue environment on mount etna. In *Isr 2018; 50th international symposium on robotics* (pp. 1–7).