NON-RECOVERING FIELD-OF-VIEW IMAGING-BASED SLAM FOR LAVA TUBES EXPLORATION

César Debeunne[1], **Alex Torres**[2], **and Damien Vivet**[1]

[1]*ISAE-SUPAERO, University of Toulouse, France* *
[2]*CNES, France* †

ABSTRACT

Visual Odometry (VO) has emerged as the go-to navigation solution for space missions, providing a reliable and efficient means of estimating the motion and state of unmanned vehicles. Recently, the exploration of potential extraterrestrial lava tubes has garnered significant attention within space agencies, presenting a captivating challenge in applied robotics. Navigating through these unique environments demands a wide Field-of-View (FoV) to efficiently catch information on the majority of the surroundings; however, the majority of rovers are equipped with narrow FoV stereo-vision configurations for direct mapping, traversability estimation and navigation. Conventional earth-based approaches to expand the FoV often prove unsuitable for space applications due to power, processing, or technical constraints. In this paper, we propose an innovative solution: indirect bi-monocular VO with a non-recovering FoV in order to make the most effective use of available camera pixels. Leveraging sliding-window optimisation techniques, we aim to overcome the inherent difficulties in accurately estimating scale in displacement and environment. Our approach paves the way for achieving robust scale-aware navigation in a non-overlapping camera configuration, opening new frontiers for space exploration methodologies in extraterrestrial lava tubes.

Key words: SLAM; Visual Odometry; Navigation.

1. INTRODUCTION

Space exploration robotic faces a significant challenge in navigating unstructured environments, as they lack the global positioning and real-time human supervision readily available on Earth. Consequently, robust navigation solutions relying on embedded sensors have become crucial. A major breakthrough in this field was the implementation of a Visual Odometry (VO) system during the Mars Exploration Rover (MER) mission, which utilized the NavCam stereo rig [GMM02]. This system provided reliable displacement estimates, overcoming the issues caused by slippage and corrupted odometers.

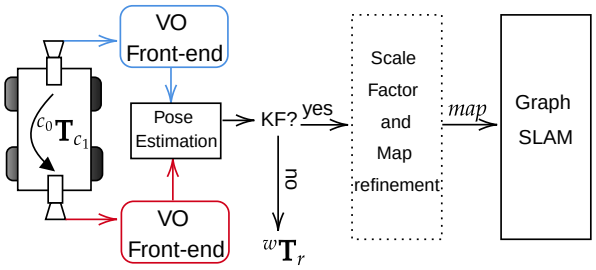


Figure 1: Our Non-Recovering FoV VO principle. A pose estimation module uses both cameras' front-end and determines if the current frame is a KeyFrame (KF) or not. If it is a KF, the scale and the landmarks between the last two KF are refined, and a local map graph optimization is performed.

The recent discovery of lava tubes on Mars and the Moon presents new challenges and opportunities. These underground structures offer protection from spatial radiation and impacts, making them potential sites for extraterrestrial human bases. Exploring these caves requires advanced navigation techniques. On Earth, cave exploration has extensively focused on Simultaneous Localization and Mapping (SLAM) relying on power-intensive Light Detection and Ranging (LiDAR) devices or offline solutions [ZB14]. Exploration of underground environments requires a wide Field-of-View (FoV) for the sensing platform often captured by rotating LiDAR. However, such solutions require last-generation embedded computers and are power-consuming.

To address these challenges for space exploration, passive sensors must be utilized. Hence, we propose a visual navigation solution for an underground context, utilizing a non-overlapping bi-monocular system. In our previous work [DVT22], a low computational cost front-end de-

*firstname.lastname@isae-supaero.fr
†firstname.lastname@cnes.fr

signed specifically for extreme lighting conditions was introduced. In this paper, our objective is to explore the expansion of the FoV to maximize information about the robot's surroundings while maintaining a robust and accurate state estimation that preserves accurate scale.

Using multiple cameras has emerged as a promising technology for robots and vehicles, offering broad FoV and high resolution. Despite their potential benefits, the current state-of-the-art SLAM systems primarily focus on monocular or stereo setups, overlooking the advantages of multi-camera configurations. The idea of utilizing non-recovering FoV camera setups has been investigated in several works in the robotic community [Ple03] [CKF+08]. These works predominantly aim to create a virtual camera from all available sensors or consist of a dual monocular/stereo configuration. In this paper, we propose a direct utilization of two monocular cameras (Figure 1) applied to space robotics to estimate the trajectory and the map at scale, offering a novel approach to leverage the benefits of multi-camera setups in space. The contributions of this paper are as follows:

- A VO designed for space exploration with two non-overlapping cameras,
- A novel approach for scale estimation with synchronized cameras using both robust initialization and on-the-fly refinement,
- The integration of observations from both cameras into a single graph-slam approach,
- A validation of our approach on both simulated and real data. This dataset is accessible online, and to our knowledge, this is the only dataset with such a camera setup available for robotic research.

2. RELATED WORK

2.1. Cave Exploration and Visual SLAM

Extensive research has been done on SLAM techniques in underground environments. Early work by Zlot *et al.* [ZB14] introduced a large-scale SLAM methodology designed to map a 17 km underground mine, employing a 2D LiDAR and a high-precision Inertial Measurement Unit (IMU). More recent developments were showcased in the DARPA Subterranean Challenge, which highlighted significant progress in multi-robot SLAM for subterranean exploration. However, the predominant reliance on LiDAR technology and high-performance embedded computers characterizes these solutions. A comprehensive survey [EBB+22] highlights the urgent need for research on cost-effective navigation solutions that integrate vision-based SLAM methodologies.

A first step toward this direction was taken by Kasper *et al.* [KMH19] that released the OIVIO dataset: a visual-inertial benchmark under extreme lighting conditions. It was recorded on a rover in dark environments like tunnels, mines and forests using onboard illumination via an LED matrix. Pure direct visual methods [EKC17] and indirect methods [CER+21] were studied, as well as Visual-Inertial Odometry (VIO) [LLB+14]. Direct methods perform direct image alignment via photometric error minimization, while indirect methods extract salient features from images to perform bundle adjustment. Overall, these methods performed well, but the VIO and the direct method seemed more robust to sudden lighting changes.

However, indirect methods have really interesting properties, especially concerning computation load, and were well studied in this context. To make these solutions more robust to poor lighting conditions, contrast enhancement techniques such as CLAHE can be applied to input images [FEM+21]. In [YYY22], several contrast enhancement techniques are applied to ORB-SLAM [MAT17] on the OIVIO dataset as well as other publicly available benchmarks. The retained solution is an improved truncated Adaptive Gamma Correction (AGC) with unsharp masking. The OIVIO dataset was also used to validate a lightweight back-end for indirect VO based on factor graph sparsification [DVTV23].

2.2. Non-overlapping field-of-view camera setup

Using a camera setup with non-overlapping FoV is interesting when it comes to state estimation and mapping. It enables a wider coverage of the surroundings of a robot. Initially, an interesting theoretical framework for a multi-camera approach is provided in [Ple03], where a camera network is represented as a single device using rays to describe pixels. This work proposes a numerical analysis of the Fisher Information Matrix of the ego-motion problem for different camera configurations. They conclude that a setup made of two cameras facing opposite directions on the same axis is the one that leverages the most uncertainties and ambiguities.

However, recovering the scale of the motion from such a setup is not as trivial as standard approaches using stereo or bi-mono with a shared FoV setup. In the approaches from the literature, the transformation between the two cameras is always known, and even this extrinsic calibration step is a challenge in itself. For instance, a method using a fixed target and camera network with a moving planar mirror is detailed in [KIFP08].

The scale recovery problem was first tackled in [CKF+08], where a method using a single point association on the second camera retrieves the scale of the motion of the first camera. The authors noticed that straight line and Ackermann motions were degenerate cases for scale estimation, which makes this problem difficult to tackle in classic VO scenarios. In [KKN+12], two monocular VO run in parallel while the scale is constantly estimated on a sliding window of keyframes in a RANSAC scheme. The 6D pose in a common frame is then retrieved by computing a weighted average of the

two VO poses by their respective covariance that is computed with the Hessian matrix of the BA. A unified approach, suitable for real-time applications, is proposed in [WK17] where multi-camera BA is performed thanks to a complete bootstrapping scheme that initializes both camera 3D positions and landmark depths.

3. METHODOLOGY

We assume that we have two synchronized cameras that do not share any part of their FoV but whose relative pose (*i.e.* the extrinsic) is known. By using the rigid transformation between the cameras, we can leverage the scale ambiguity and improve the consistency of the VO. We note ${}^{c_0}\mathbf{T}_{c_0'}, {}^{c_1}\mathbf{T}_{c_1'} \in SE(3)$ the motions of the two cameras between two KeyFrames (KF) and their extrinsic calibration is given by ${}^{c_0}\mathbf{T}_{c_1}$.

The overall system is described in Figure 1. The VO front-end, as well as the graph SLAM back-end, are detailed in [DVT22], this section describes what is specific for a VO with non-overlapping cameras. These are the bootstrapping phase, the pose estimation module and the scale plus map refinement module.

3.1. Initialization of the scale

In the beginning, the camera c_0 exhibits a motion ${}^{c_0}\mathbf{T}_{c_0'} = [{}^{c_0}\mathbf{R}_{c_0'} | \lambda \mathbf{t}_{c_0}^{c_0'}]$ whose scale λ is unknown. This motion can be computed using epipolar geometry [HZ03] and leads to an ambiguous expression of the second camera motion: ${}^{c_1}\mathbf{T}_{c_1'} = {}^{c_0}\mathbf{T}_{c_1}^{-1} {}^{c_0}\mathbf{T}_{c_0'} {}^{c_0}\mathbf{T}_{c_1}$. Derivations conducted in [CKF+08] return a formula for the scale λ derived from the essential matrix of the views of the second camera. Considering the two corresponding homogeneous points $\mathbf{x}' \longleftrightarrow \mathbf{x}$ observed on c_1, the scale factor is given by:

$$\lambda = -\frac{\mathbf{x}' \left({}^{c_0}\mathbf{R}_{c_1}^T \left[{}^{c_0}\mathbf{R}_{c_0'} \mathbf{t}_{c_0}^{c_1} - \mathbf{t}_{c_0}^{c_1} \right]_\times {}^{c_0}\mathbf{R}_{c_0'} {}^{c_0}\mathbf{R}_{c_1} \right) \mathbf{x}}{\mathbf{x}' \left({}^{c_0}\mathbf{R}_{c_1}^T \left[\mathbf{t}_{c_0}^{c_0'} \right]_\times {}^{c_0}\mathbf{R}_{c_0'} {}^{c_0}\mathbf{R}_{c_1} \right) \mathbf{x}}. \quad (1)$$

We have implemented this equation in a single-point RANSAC scheme to fix the scale ambiguity of these motions. Degenerate motions, which are detailed in [CKF+08], are detected and ignored. The process is restarted until a proper motion is performed: the VO can only begin once the scale is initialized. After initialization, the scale and the map are refined with a "scale-only" bundle adjustment.

3.2. Pose Estimation

The VO front-end performs tracking of 2D keypoints that can have 3D coordinates if they are from the local map.

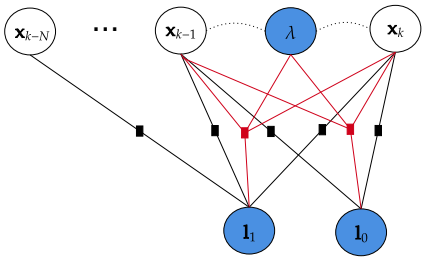


Figure 2: A factor graph representation of the scale and map refinement problem. In circles are the variables, and the plain lines represent the residuals linking variables. The blue variables are the ones that are optimized, and in red are the visual residuals that include the scale of the motion between $\mathbf{x}_{k-1}$ and $\mathbf{x}_k$.

2D-3D associations enable the estimation of the 3D pose of the camera with a Perspective-n-Points (PnP) algorithm. Here we use, for both cameras, the variant P3P [KSS11] in a RANSAC scheme to calculate ${}^{w}\mathbf{T}_{c_0}$, ${}^{w}\mathbf{T}_{c_1}$ and their respective covariances $\Sigma_{c_0}, \Sigma_{c_1}$.

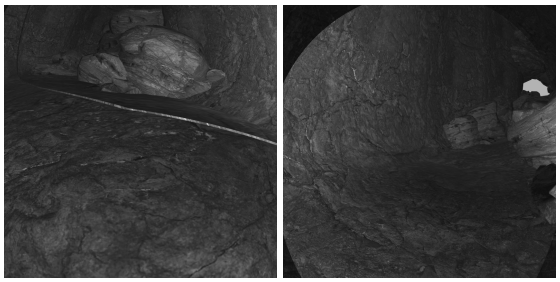
These two poses are then fused to compute the current poses of the robot in a similar way as in [KKN+12]. For both cameras, we compute the associated robot pose using the extrinsic ${}^{w}\mathbf{T}_{r_0} = {}^{w}\mathbf{T}_{c_0} {}^{c_0}\mathbf{T}_r$ and ${}^{w}\mathbf{T}_{r_1} = {}^{w}\mathbf{T}_{c_1} {}^{c_1}\mathbf{T}_r$. Then we compute the difference between the two poses on a tangent space as $\delta\mathbf{T} = {}^{w}\mathbf{T}_{r_0} \ominus {}^{w}\mathbf{T}_{r_1} \in \mathbb{R}^6$. The covariances computed with the pose estimator give us the weight $W = \Sigma_{c_0}(\Sigma_{c_0} + \Sigma_{c_1})^{-1}$ that is used to fuse the poses as:

$${}^{w}\mathbf{T}_r = {}^{w}\mathbf{T}_{r_0} \oplus W\delta\mathbf{T}. \quad (2)$$

The $\ominus$ and $\oplus$ operators come from the $\mathrm{SO}(3) \times \mathbb{R}^3$ manifold that is used to parameterize the poses in our framework; more details are given in [SDA18].

3.3. Scale and Map Refinement

As in [DVT22], when a KF is declared, the 2D points tracked since the last KF are triangulated and become part of the SLAM problem. Usually, a reprojection error minimization is performed on these new landmarks to refine their 3D coordinates. But here, to guarantee a consistent scale, we also add the scale of the motion of a camera in this problem. At timestep i, we note the up-to-scale motion of camera c_0 between the two last KFs ${}^{c_0}\mathbf{T}_{c_0'}(\lambda) = \left[{}^{c_0}\mathbf{R}_{c_0'} | \lambda \mathbf{t}_{c_0}^{c_0'} \right]$ and, to remain general, we note $e_V \left({}^{w}\mathbf{T}_c, \mathbf{l}_w^j \right)$ the visual error of a landmark located at $\mathbf{l}_w^j \in \mathbb{R}^3$ on a camera at pose ${}^{w}\mathbf{T}_c$. The set of landmarks observed by c_0 is noted L_0, respectively L_1. For a given landmark $\mathbf{l}_w^j$, we note $\mathscr{C}_j$ the set of cameras from which it is observed. The refinement consists of finding the solu-



(a) (b)

Figure 3: Two synchronous images from our simulated fisheye cameras in a cave environment with self-illumination.

tion to the following problem:

$$\underset{\lambda, L_0, L_1}{\operatorname{argmin}} \sum_{\mathbf{l}_w^j \in L_0} \left[||e_V({}^w\mathbf{T}_{c_0}\, {}^{c_0}\mathbf{T}_{c_0'}(\lambda), \mathbf{l}_w^j)||^{\rho}_{\Sigma_V} + \sum_{c \in \mathscr{C}_j} ||e_V({}^w\mathbf{T}_c, \mathbf{l}_w^j)||^{\rho}_{\Sigma_V} \right] + \sum_{\mathbf{l}_w^j \in L_1} \left[||e_V({}^w\mathbf{T}_{c_0}\, {}^{c_0}\mathbf{T}_{c_0'}(\lambda)\, {}^{c_0}\mathbf{T}_{c_1}, \mathbf{l}_w^j)||^{\rho}_{\Sigma_V} + \sum_{c \in \mathscr{C}_j} ||e_V({}^w\mathbf{T}_c, \mathbf{l}_w^j)||^{\rho}_{\Sigma_V} \right]. \tag{3}$$

Taking into account both sets of landmarks from c_0 and c_1, the optimal scale respects the rigid transformation between the two cameras. The intuition behind this minimization is that this will force the scale to remain stable if there are many observations from the current local map. However, when a lot of landmarks are lost, there will be less error terms from the local map. This will perform a "new initialization" and correct the scale that may have gone wrong. A factor graph representation of this problem is represented in figure 2 to highlight the residuals involved and the variables that they link. After this step, the landmarks with visual errors that do not pass a ξ^2 test are removed from the problem. The robust Huber loss function ρ is used to mitigate the presence of these outliers, and the covariance Σ_V for visual observations is set to the identity.

4. EXPERIMENTAL RESULTS

This section introduces experimental results that allow us to validate our method on a great variety of scenarios[1]. Our algorithm is developed in C++, using the middleware ROS2 to provide communication with our sensor suite and the library CERES for non-linear optimization. All experiments have been performed on a desktop station equipped with an Intel Core i7, 3.2 GHz clock rate, using CPU only.

4.1. Simulated Data

We have first tested this approach on data from the simulator Gazebo. The dataset is recorded with fisheye cameras mounted on a rover: one is oriented toward the front, and the other to the rear. The first scenario is an easy one on a planetary surface with good lighting conditions. The second scenario is in a cave world inspired by the virtual DARPA Subterranean challenge [KKV+20] using onboard illumination for both cameras, which makes it more challenging. Both trajectories start with a 90-degree turn that ensures that the scale will be properly initialized. An extract from our dataset is shown in figure 3.

[1]The dataset used in this paper is available at https://doi.org/10.34849/SEVFJB

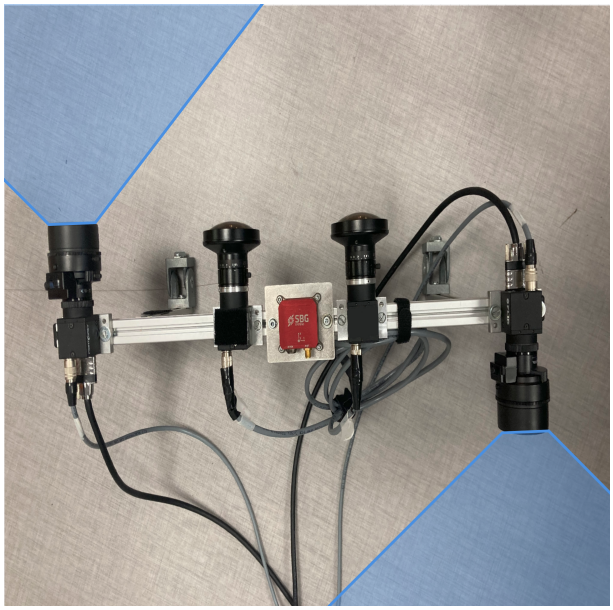


Figure 4: This is our camera setup. The two FLIR cameras, whose FoV is drawn in blue, are facing opposite directions. The extrinsic was computed using the SBG IMU, and the two fisheye cameras are not used in the following experiments.

We have tested three different methods: our full method, our method without scale refinement, and our VO in mono mode with a scale initialized as in [CKF+08]. Three metrics were computed to compare these methods. The Absolute Trajectory Error (ATE) in meters, which is the error wrt. the aligned ground truth trajectory. The Relative Angular Error (RAE) in degree is for relative angular displacement, and the Scale Error (SE) without units evaluates the ratio between the norm of the translational displacements:

$$SE = \frac{1}{N_{KF}} \sum_{k=1}^{N_{KF}} \left| 1 - \frac{||\mathbf{t}_{VO}^k||^2}{||\mathbf{t}_{gt}^k||^2} \right|, \tag{4}$$

where $\mathbf{t}_{VO}^k$ and $\mathbf{t}_{gt}^k$ denote the translation between two successive KFs estimated respectively by the VO and the ground truth system. The alignment between the ground truth and the estimated trajectories is recovered using 6-DoF optimization; the scale is never rectified. The results of this study are summarized in Table 1.

Our full method provides the best results in both ATE and SE on the two simulated trajectories. The scale refinement module enables a better scale of robot motion that leads to better overall accuracy. The mono method is per-

Scenario	Our method			Our method w/o scale refinement			Mono VO		
	ATE(m)	RAE(degree)	SE	ATE(m)	RAE(degree)	SE	ATE(m)	RAE(degree)	SE
Planetary	**0.01**	**0.28**	**0.13**	0.03	0.30	**0.13**	0.09	0.31	0.31
Cave	**0.02**	0.12	**0.08**	0.10	**0.11**	0.09	0.47	0.12	0.11
Chariot 1	**0.04**	**0.68**	**0.09**	0.20	0.69	0.17	0.34	**0.68**	0.28
Chariot 2	**0.05**	0.76	**0.08**	0.17	**0.71**	0.13	0.37	0.75	0.30

Table 1: Performances on both simulated and indoor datasets: the best results for each metric are displayed in bold.

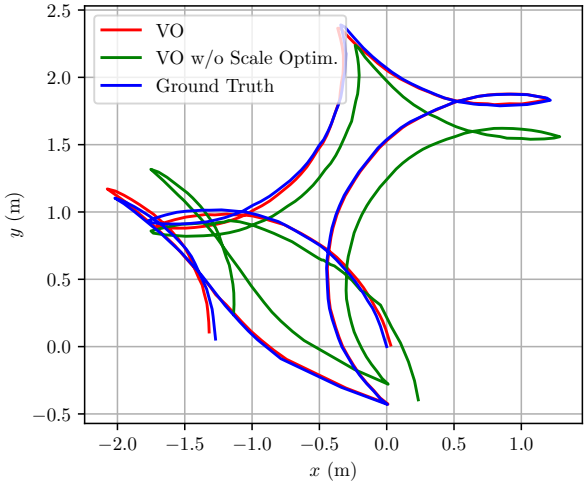


Figure 5: 2D comparison between our full system, our system without scale optimization and the ground truth on the *chariot1* trajectory.

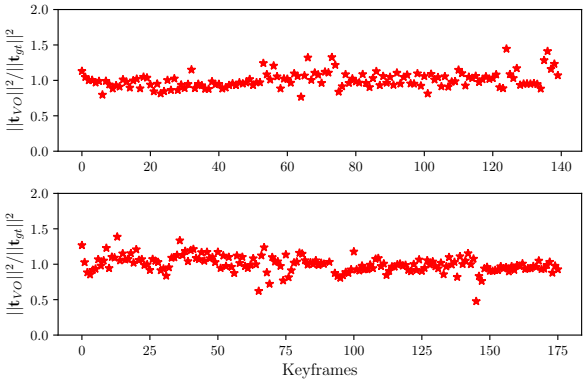


Figure 6: Plot of the ratio of the displacement between successive KFs evaluated by the VO and by the Motion Capture system. The top is for the *chariot1* scenario and the bottom for the *chariot2* scenario.

forming way worse on ATE and SE than the two others, the scale drift is easily accumulated using a single camera and performing BA on two cameras (even without scale refinement) helps a lot to reduce this drift. However, the RAE metric seems to be similar for all the approaches. This is logical, a camera can be considered as a bearing-only sensor whose strength is orientation estimation. Using additional cameras improves only translational motion estimation.

4.2. Real Data

To validate our method on real data that may suffer from calibration noise, we mounted an experimental bench with two FLIR Blackfly cameras facing opposite directions on our already existing VIO setup as shown on figure 4. We performed extrinsic calibration using the fixed IMU on the bench, and the Kalibr calibration toolbox [FRS13]. This gives us the two following visual-inertial calibrations: ${}^{\text{imu}}\mathbf{T}_{c_0}$ and ${}^{\text{imu}}\mathbf{T}_{c_1}$. With a simple transformation chain, we obtain the extrinsic of our non-overlapping FoV configuration:

$$ {}^{c_0}\mathbf{T}_{c_1} = {}^{\text{imu}}\mathbf{T}_{c_0}^{-1}\,{}^{\text{imu}}\mathbf{T}_{c_1}. \tag{5} $$

4.2.1. Indoor dataset

We first recorded a data set in an indoor environment equipped with a motion capture system at ISAE-SUPAERO, Toulouse, France. This provides a precise pose measurement at a high rate that we consider a ground truth.

The same conclusions concerning the comparison between the three methods can be deduced from the table 1 on this real dataset. A comparison between our full system with and without scale refinement is shown in figure 5: without scale optimization, the alignment with the ground truth is impossible due to scale inconsistencies in the trajectory. One can notice that our full system reaches decent accuracy on a real setup for trajectories that are respectively 14m and 20m long. Moreover, as shown in figure 6, the scale remains consistent all along the trajectory: no drift can be noticed, and the ratio is around the optimal value of one.

4.2.2. Outdoor dataset

To test our method on a larger scale, two other scenarios were recorded outdoor on a rover on the campus of ISAE-SUPAERO. A trajectory named *forest* is performed under the trees and on bumpy ground, the illumination conditions are poor, and the scene is highly unstructured. The other trajectory, *square*, is recorded on the road around a building of the campus: it is less challenging in terms of

	Length (m)	Duration (s)	Drift (%)
Forest	108	113	2.1
Square	236	205	0.5

Table 2: Performance of our system and description of the outdoor sequences

	Front-end	Pose estimator	Scale refinement	BA	**total**
t(ms)	18.6	1.88	8.01	7.52	24.5

Table 3: Timing results on our Desktop computer

motion, but it exhibits poorly textured scenes due to the tarmac and the sky. The ground truth was obtained via a differential GNSS system that offers positioning precision at the centimeter level. The metric evaluated is the drift, in percentage, that we define as the ATE normalized by the length of the trajectory. A summary of the performance of our system is provided in table 2. The drift is around a percent on large-scale scenarios which is reasonable for exploration scenarios: it is less than the 3 % objective stated in [GMM02]. However, we have noticed that in an extreme case of featureless scenes, our algorithm was relying only on one of the two cameras for motion estimation, which have led to a wrong scale estimate on several occasion during the *forest* scenario. In the *square* scenario, the scale was consistent all over the trajectory even with the presence of long straight lines that are degenerated cases for scale recovery; this can be observed on figure 7. Moreover, loops were closed in these scenarios, but our system doesn't handle loop closure yet: this may have drastically corrected the drift and is left for future work.

4.3. Run time analysis

For spatial applications, the computational burden is a real issue both because of light hardware and limited energy resources. This algorithm was designed to limit calculations on the front-end [DVT22], however, this camera configuration leads to doubling the size of the map in comparison to mono or bi-mono setup. Moreover, the two VO front-ends are not parallelized, which also doubles the run time for feature extraction and tracking.

For this study, we have used the same configuration as in the previous experiments: 150 keypoints are maintained in each image, 10 KFs are kept in the sliding window, and a KF is voted if the average parallax goes over 3 degrees. The results for each algorithmic step are displayed on table 3. Overall, our algorithm runs at 40Hz, which makes it suitable for real-time, but for the reasons detailed before, it is slower than our bi-mono VO that can reach 100 Hz.

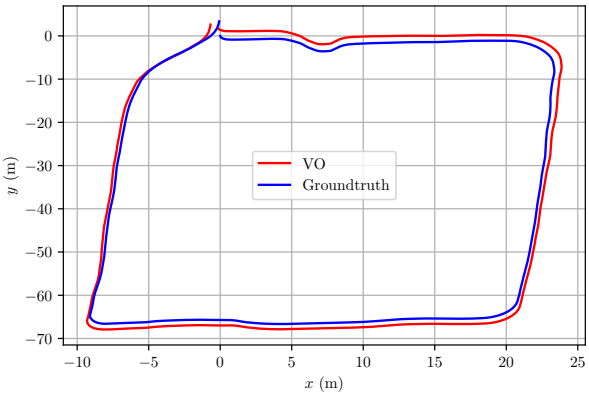


Figure 7: Comparison between the VO trajectory and the ground truth on the *square* scenario.

5. CONCLUSION AND PERSPECTIVES

This paper has introduced a method to perform scale estimation and VO using a pair of non-overlapping FoV cameras. This work was motivated to enable safe navigation for extraterrestrial Lava Tubes exploration. A complete study on simulated, indoor and outdoor data was conducted to demonstrate the performances and the limitations of our method on a large spectrum of scenarios. However, apart from simulated data, this system was not tested on a real, large-scale, underground and self-illuminated scenario. We would like to conduct a more complete recording campaign to produce a dataset similar to [KMH19] with our camera setup. We also want to extend our system, that is limited to state estimation, by implementing a surface reconstruction module to produce traversability information.

ACKNOWLEDGEMENT

This work was supported by the CNES and Occitanie Region. We would like to thank sincerely Benoit Priot and Corentin Chauffaut for their help during the experiments.

REFERENCES

[CER+21] Carlos Campos, Richard Elvira, Juan J. Gómez Rodríguez, José M. M. Montiel, and Juan D. Tardós. ORB-SLAM3: An accurate open-source library for visual, visual–inertial, and multimap slam. *IEEE Transactions on Robotics*, 37(6):1874–1890, 2021.

[CKF+08] Brian Clipp, Jae-Hak Kim, Jan-Michael Frahm, Marc Pollefeys, and Richard Hartley. Robust 6dof motion estimation for non-overlapping, multi-camera systems. In *2008 IEEE Workshop on Applications of Computer Vision*, pages 1–8, 2008.

[DVT22] César Debeunne, Damien Vivet, and Alex Torres. Design of a bi-monocular Visual Odometry System for Lava Tubes exploration. In *PNARUDE Workshop 2022 IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2022.

[DVTV23] César Debeunne, Joan Vallvé, Alex Torres, and Damien Vivet. Fast bi-monocular Visual Odometry using Factor Graph Sparsification. In *2023 IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2023.

[EBB+22] Kamak Ebadi, Lukas Bernreiter, Harel Biggie, Gavin Catt, Yun Chang, Arghya Chatterjee, Christopher E. Denniston, Simon-Pierre Deschênes, Kyle Harlow, Shehryar Khattak, Lucas Nogueira, Matteo Palieri, Pavel Petráček, Matěj Petrlík, Andrzej Reinke, Vít Krátký, Shibo Zhao, Ali-akbar Agha-mohammadi, Kostas Alexis, Christoffer Heckman, Kasra Khosoussi, Navinda Kottege, Benjamin Morrell, Marco Hutter, Fred Pauling, François Pomerleau, Martin Saska, Sebastian Scherer, Roland Siegwart, Jason L. Williams, and Luca Carlone. Present and future of SLAM in extreme underground environments, 2022.

[EKC17] J. Engel, V. Koltun, and D. Cremers. Direct sparse odometry. In *IEEE Transactions on Pattern Analysis and Machine Intelligence*, volume 40, pages 611–625, 2017.

[FEM+21] Maxime Ferrera, Alexandre Eudes, Julien Moras, Martial Sanfourche, and Guy Le Besnerais. Ov^{2}slam: A fully online and versatile visual slam for real-time applications. *IEEE Robotics and Automation Letters*, 6(2):1399–1406, 2021.

[FRS13] Paul Furgale, Joern Rehder, and Roland Siegwart. Unified temporal and spatial calibration for multi-sensor systems. In *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 1280–1286, 2013.

[GMM02] S.B. Goldberg, M.W. Maimone, and L. Matthies. Stereo vision and rover navigation software for planetary exploration. In *Proceedings, IEEE Aerospace Conference*, volume 5, pages 2025–2036, 2002.

[HZ03] Richard Hartley and Andrew Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, New York, NY, USA, 2003.

[KIFP08] Ram Krishan Kumar, Adrian Ilie, Jan-Michael Frahm, and Marc Pollefeys. Simple calibration of non-overlapping cameras with a mirror. In *2008 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–7, 2008.

[KKN+12] Tim Kazik, Laurent Kneip, Janosch Nikolic, Marc Pollefeys, and Roland Siegwart. Real-time 6d stereo visual odometry with non-overlapping fields of view. In *2012 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1529–1536, 2012.

[KKV+20] Anton Koval, Christoforos Kanellakis, Emil Vidmark, Jakub Haluska, and George Nikolakopoulos. A subterranean virtual cave world for gazebo based on the darpa subt challenge. *ArXiv*, abs/2004.08452, 2020.

[KMH19] Mike Kasper, Steve McGuire, and Christoffer Heckman. A Benchmark for Visual-Inertial Odometry Systems Employing Onboard Illumination. In *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2019.

[KSS11] Laurent Kneip, Davide Scaramuzza, and Roland Siegwart. A novel parametrization of the perspective-three-point problem for a direct computation of absolute camera position and orientation. In *Conference on Computer Vision and Pattern Recognition*, pages 2969–2976, 2011.

[LLB+14] Stefan Leutenegger, Simon Lynen, Michael Bosse, Roland Siegwart, and Paul Furgale. Keyframe-based visual-inertial odometry using nonlinear optimization. *The International Journal of Robotics Research*, 34, 02 2014.

[MAT17] Raul Mur-Artal and Juan D. Tardos. ORB-SLAM2: An open-source SLAM system for monocular, stereo, and RGB-d cameras. *IEEE Transactions on Robotics*, 33(5):1255–1262, 2017.

[Ple03] Robert Pless. Using many cameras as one. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 2, pages II–587. IEEE, 2003.

[SDA18] Joan Solà, Jeremie Deray, and Dinesh Atchuthan. A micro lie theory for state estimation in robotics, 2018.

[WK17] Yifu Wang and Laurent Kneip. On scale initialization in non-overlapping multi-perspective visual odometry. pages 144–157, 10 2017.

[YYY22] Leijian Yu, Erfu Yang, and Beiya Yang. Afe-orb-slam: Robust monocular vslam based on adaptive fast threshold and image enhancement for complex lighting environments. *Journal of Intelligent and Robotic Systems*, 105, 05 2022.

[ZB14] Robert Zlot and Michael Bosse. Efficient large-scale 3d mobile mapping and surface reconstruction of an underground mine. In *Field and service robotics*, pages 479–493, 2014.