Precise pose estimation of the NASA Mars 2020 Perseverance rover through a stereo-vision based approach

Simone Andolfo [1*], Flavio Petricca[1], Antonio Genova[1]

[1] Department of Mechanical and Aerospace Engineering, Sapienza University of Rome, Rome, Italy

*Corresponding author. Email: simone.andolfo@uniroma1.it

ABSTRACT

Visual Odometry (VO) is a fundamental technique to enhance the navigation capabilities of planetary exploration rovers. By processing the images acquired during the motion, VO methods provide estimates of the relative position and attitude between navigation steps with the detection and tracking of 2D image-keypoints. This method allows to mitigate trajectory inconsistencies associated with slippage conditions resulting from dead-reckoning techniques. We present here an independent analysis of the high-resolution stereo images of the NASA Mars 2020 Perseverance rover to retrieve its accurate localization on sols 65, 66, 72, and 120. The stereo pairs are processed by using a 3D-to-3D stereo-VO approach that is based on consolidated techniques and accounts for the main nonlinear optical effects characterizing real cameras. The algorithm is first validated through the analysis of rectified stereo images acquired by the NASA Mars Exploration Rover (MER) Opportunity, and then applied to the determination of Perseverance's path. The results suggest that our reconstructed path is consistent with the telemetered trajectory, which was directly retrieved onboard the rover's system. The estimated pose is in full agreement with the archived rover's position and attitude after short navigation steps. Significant differences (~10-30 cm) between our reconstructed and telemetered trajectories are observed when Perseverance travelled distances larger than 1 m between the acquisition of stereo pairs.

KEYWORDS: visual odometry; rovers; planetary exploration; space robotics; stereo vision; localization; computer vision.

1. INTRODUCTION

In July 1997, as part of NASA's Mars Pathfinder mission, the Sojourner rover became the first vehicle to drive on the planet Mars. During its 83-days mission, Sojourner explored the area near its landing site called Ares Vallis, travelling ~100 meters while capturing images of the Martian landscape. The acquired stereo pairs were processed in combination with light-striper sensors to detect hazards (*e.g.*, rocks, depressions) in the rover's proximity, supporting navigation operations (Mishkin et al, 1998). However, Sojourner's localization software did not include information from the acquired images, and the rover's position and attitude (*i.e.*, pose) were updated through dead-reckoning by combining Inertial Measurement Units (IMUs) and wheel odometry (WO) measurements.

Dead-reckoning represents the basic method to update the pose of rovers exploring planetary environments. This method is affected by significant errors associated with the slippage that accumulate over time. To compensate for dead-reckoning errors, Visual Odometry (VO) techniques enables highly accurate pose estimates of moving assets by tracking fiducial points of a scene observed by the onboard cameras. VO was first used for planetary applications by the Mars

Exploration Rovers (MER) Spirit and Opportunity (Biesiadecki & Maimone, 2006; Matthies et al., 2007; Maimone et al., 2007). The MER-VO algorithm was based on the determination of 3D coordinates of selected keypoints after stereo-matching of left- and right-eye images through correlation methods. These keypoints were then tracked in the new stereo pairs, and a maximum-likelihood filter was used in a 3D-to-3D pose estimation problem (Matthies & Shafer, 1987). The MER-VO algorithm enabled accurate pose estimates by measuring position variations as small as 2 mm even on steep terrains (*e.g.*, slopes >30°) (Maimone et al., 2007). However, limited computational resources onboard both rovers were not well-suited for continuous Guidance, Navigation, & Control (GNC) operations with the support of VO. The image processing algorithm required 2-3 minutes for each drive step, dramatically limiting the rover's speed to ~8% of the maximum speed (*i.e.*, ~120 m/hour in "blind" drive sessions based on the execution of the navigation commands sent by the ground operations team). Autonomous safe exploration of rough terrains was also limited since the VO localization and autonomous hazard detection software were barely used simultaneously.

To enhance the response time of the image processing scheme for the NASA Mars Science Laboratory (MSL) Curiosity rover, a refined stereo correlation algorithm and an iterative image pyramid scheme were included in the VO algorithm (Johnson et al., 2008). By going through all VO stages, from feature selection to motion estimate, iteratively at each level of the image pyramid, the MSL-VO algorithm allows to constrain the search of tracked features in the finer resolution images. This iterative approach yields significant computational time savings, and false feature tracking is limited once the bottom of the image pyramid is reached. This algorithm then led to obtain a motion estimate in 47 seconds on average, successfully processing 99.55% of the taken drive steps during the first seven years of the mission (Rankin et al., 2020). VO also played a crucial role in preserving the rover's safety, as demonstrated during the MSL path replanning towards Mount Sharp after the detection of an unexpected high slippage of the wheels over the rippled sand of the Hidden Valley.

The NASA's MER and MSL missions paved the way to the accurate localization of planetary rovers on heterogeneous and demanding terrains through VO algorithms. This solid technique will be used by future and current missions, including the CNSA lunar Yutu-2 rover (Ma et al., 2020), the NASA Mars 2020 rover Perseverance, which landed on Mars in February 2021, and the ESA-Roscosmos ExoMars rover (Townson et al., 2018; Winter et al., 2015), which was planned to be launched in 2022. Perseverance is currently exploring the Jezero crater searching for signs of ancient life and investigating the geological evolution of the planet (Farley et al., 2020). The rover's navigation system represents the state of the art in planetary surfaces exploration. Compared to the previous rovers, Perseverance can move across the Martian surface more autonomously, and the visual input from the onboard navigation cameras (NavCams) are processed to continuously replan its trajectory during the motion, without a stop-and-go approach. Perseverance hosts onboard dedicated hardware to carry out demanding computer vision tasks for a safe path planning and an accurate localization based on VO (Verma, 2020; Verma et al., 2022).

In addition to support autonomous scientific operations in difficult planetary environments, advanced vision-based localization systems have been employed in a wide range of terrestrial activities carried out on ground (Nistér et al., 2004; Howard, 2008; Scaramuzza et al., 2009), and in challenging aerial (Kim et al., 2019) and underwater scenarios (Ferrera et al., 2019; Teixeira et al., 2020). These applications will pave the way to future exploration missions to remote areas in the Solar System, including the icy moons' oceans, and dense atmospheric environments (Witte et al., 2019). Although stereo-VO represents the baseline for planetary applications, the use of single omnidirectional (Corke et al., 2004) and monocular cameras has been investigated to support flying robots operations (*e.g.*, the Ingenuity helicopter; Wudenka et al., 2021), to estimate the motion of a hopping rover on irregular asteroid surfaces (So et al., 2011), and to measure the rover's slippage on loose terrains (Gonzalez & Iagnemma, 2018).

In this paper, we present the results concerning an alternative and independent reconstruction of Perseverance path through a stereo-VO algorithm based on the 3D-to-3D formulation (Matthies & Shafer, 1987), which processes images captured by the rover's NavCams (Maki et al., 2020). The camera model adopted in this study is presented in Section 2, and a step-by-step description of the VO algorithm is then discussed in Section 3. A validation of the method is provided in Section 4 by retrieving pose estimates of Opportunity through the processing of rectified stereo NavCams images. Section 5 is focused on the reconstruction of Perseverance's path that is obtained by analyzing raw stereo NavCams image pairs with the proposed VO algorithm.

The processing of data acquired by current and past planetary rover missions represents a significant testbed to assess the performances of image-based localization systems. This study provides accurate information on the pose estimation precision that can be attained with an autonomous navigation system, which is currently under development by our research group to support rover prototypes' operations on unprepared terrains.

2. ROVER AND CAMERA MODELING

To define the parameters that will be adjusted in the filter, we present the adopted models of the camera and the rover in Sections 2.1 and 2.2, respectively. A thorough description of the camera modeling is important to correctly convert 2D image-points into 3D world-points forward and backward.

2.1 CAHVORE Camera Model

Images acquired by real cameras are affected by nonlinear effects (*e.g.,* optical distortion) that are not accounted for by the pinhole camera model, which adopts an undistorted perspective projection (Young, 1971). To accurately describe the acquisition geometry of wide-angle cameras employing fisheye lenses, refined camera models have been developed, including the Brown model (Brown, 1971), the Kannala-Brandt model (Kannala & Brandt, 2006), and the CAHVORE model (Gennery, 2001; Gennery, 2006). The latter is currently adopted by the engineering cameras of NASA planetary rovers.

Compared to other camera models, the CAHVORE model employs more parameters that enable a refined modeling of the radial optical distortion, by allowing for the possibility for the optical axis to be not exactly perpendicular to the camera sensor plane. In general, radial distortion is described by a polynomial that gives the departure of the off-axis coordinate from its ideal value as a function of the off-axis coordinate. The off-axis coordinate is usually expressed in terms of image coordinates (Brown, 1971; Kannala & Brandt, 2006) rather than to be defined relative to the lens optical axis, and this implies that the optical axis $\boldsymbol{O}$ is assumed to be parallel to the sensor plane's normal $\boldsymbol{A}$. Although real camera lenses are manufactured so that this assumption holds, the two vectors may be slightly misaligned, and the CAHVORE model enables to account for this effect. Furthermore, the CAHVORE model accounts for the displacement of the entrance pupil along the optical axis (Fasogbon & Aksu, 2019) that is modeled as a function on the off-axis angle of the incoming light rays. This nonlinear effect is usually ignored in camera calibration since it is small, but can be significant for wide field-of-view (FOV) cameras. For example, for the hazard cameras onboard the MER rovers, the forward entrance pupil shift can be as high as ~7 mm, leading to an error of ~4° for objects as close as 10 cm (Gennery, 2006).

The CAHVORE model efficiently describes the acquisition geometry of wide-angle cameras through a set of seven 3-dimensional vectors, which are used to define the pose of the camera and the camera intrinsic parameters, and to model the nonlinear optical effects. Each letter of the acronym CAHVORE is associated with one of these vectors, which are detailed hereafter. The *camera vector* $\boldsymbol{C}$ defines the nominal 3D location of the entrance pupil. The *axis vector* $\boldsymbol{A}$ is a unit vector orthogonal to the image plane and departs from the entrance pupil $\boldsymbol{C}$ pointing outwards. The vectors $\boldsymbol{H}$ and $\boldsymbol{V}$ are the *horizontal vector* and the *vertical vector*, respectively. $\boldsymbol{H}$ and $\boldsymbol{V}$ are combined with $\boldsymbol{A}$ to determine the image-coordinates of the camera principal point. Their projections onto the image plane ($\boldsymbol{H'}$ and $\boldsymbol{V'}$) provide vectors that are almost aligned with the image rows and columns, respectively (Di & Li, 2004). The vectors $\boldsymbol{O}$ and $\boldsymbol{R}$ are the *optical vector* and the *radial vector*, respectively. They are jointly used to model optical radial distortion. The radial vector collects the even-order coefficients of a 4-degree polynomial used to compute the displacement of the 3D points in a direction orthogonal to $\boldsymbol{O}$, which compensates the lens curvature (Gennery, 2006). The definition of distinct vectors $\boldsymbol{O}$ and $\boldsymbol{A}$ enables to account for the non-orthogonality of the image plane with respect to the optical axis; for ideal lenses or rectified images, the vectors $\boldsymbol{O}$ and $\boldsymbol{A}$ are parallel. $\boldsymbol{E}$ is the *entrance vector* collecting the even-order coefficients of a 4-degree polynomial used to model the displacement of the entrance pupil along the optical axis $\boldsymbol{O}$. The adjusted position of the entrance pupil $\boldsymbol{C'}$ is defined accordingly to:

$$\boldsymbol{C'} = \boldsymbol{C} + s\boldsymbol{O} \tag{1}$$

where $s = s(\alpha,\ \boldsymbol{E})$, and α is the off-axis angle between the incoming viewing ray and the camera optical axis (Gennery, 2006). Incoming rays aligned with the optical axis ($\alpha = 0$) produce a null displacement of the entrance pupil.

The cameras onboard NASA planetary rovers have been accurately calibrated before flight through a metrology-dependent approach that uses precisely measured dot-target positions relative to the cameras and solves for the CAHVORE parameters only. This represents a major difference with respect to standard calibration procedures that employ a pure-photogrammetric approach that jointly solves for dot-targets locations and camera parameters in a single bundle-adjustment (Hayes et al., 2021). A detailed documentation of the on-ground calibration activities reports the best-estimated CAHVORE parameters for each camera, which are defined for a specific pose of the camera with respect to the rover navigation frame. As the cameras change the pointing direction, the related 3D vectors $\boldsymbol{C}$, $\boldsymbol{A}$, $\boldsymbol{H}$, $\boldsymbol{V}$ and $\boldsymbol{O}$ are updated through kinematical equations accounting for the actual azimuth and elevation angles of the mast (Ruoff et al., 2021). The vectors $\boldsymbol{R}$ and $\boldsymbol{E}$ are fixed depending on the lens characteristics only. The updated CAHVORE parameters are then referred to the rover navigation frame, and their values are reported in the image metadata.

The VO navigation software of the NASA MER and MSL rovers processed rectified images, *i.e.*, images projected in a theoretical and distortion-free stereo setup geometry (Ruoff et al., 2021), obtained through a preprocessing of the raw images. The rectified images can be described by a simplified CAHV model (Yakimovsky & Cunningham, 1978), equivalent to a pinhole camera model (Di & Li, 2004), and are epipolar-aligned. Therefore, corresponding pixels in the left and right images can be searched for about the *horizontal* epipolar line, reducing the chance of wrong matches.

Perseverance navigation software was conceived to be independent from the rectification of the acquired images, and corresponding pixels are searched about the epipolar curve. This main change with respect to the previous missions is associated with the enhanced and upgraded design of the NavCams. Compared to MER and MSL NavCams (image detector size: 1024×1024 pixels; field of view: 45° × 45°), the Perseverance NavCams have a wider 90° × 70° field of view and acquire 20×

higher-resolution images (image detector size: 5120×3840 pixels). However, to deal with limited onboard memory resources, Perseverance flight software, inherited from MSL, processes only tiles of the full image (Ruoff et al., 2021). The maximum size of a readable tile is 1280×960 pixels, and 16 tiles are then required to read out full-resolution (1× downsampling) images; at 2× downsampling, 4 tiles are required; at 4× or 8× downsampling, only 1 tile is required, and the entire image can be read at once. These multiple image acquisition modes result in a more complex rectification procedure, depending on the actual tiling and downsampling parameters. To analyze non-rectified images, our VO algorithm accounts for the full nonlinear CAHVORE camera model.

2.2 Parametrization of the Rover's Motion

The single rover's drive step is assumed to be a 6 degrees of freedom (DOF) rigid rototranslation defined by the translation vector $\boldsymbol{\tau}$ and the rotation matrix ${}^{\{A\}}_{\{B\}}\mathbf{R}$. The motion parameters $\boldsymbol{\tau}$ and ${}^{\{A\}}_{\{B\}}\mathbf{R}$ define the (4×4) transformation matrix ${}^{\{A\}}_{\{B\}}\mathbf{T}$ from $(\boldsymbol{O}_{\mathrm{B}},\{\mathrm{B}\})$ to $(\boldsymbol{O}_{\mathrm{A}},\{\mathrm{A}\})$, which denote the rover navigation frame before and after the motion step, respectively. In this work, the rotation is expressed through the rotation vector $\boldsymbol{\Theta}$ that consists of the yaw-pitch-roll Bryant angles.

To adjust the rover's motion parameters, the VO algorithm takes as input two stereo pairs, acquired at the beginning and at the end of the drive step, and 3D points triangulated from both stereo pairs. The motion equation relates the 3D coordinates of a world-point $\boldsymbol{\mathcal{P}}$ observed before ($\boldsymbol{P}^{\{\mathrm{B}\}}$) and after ($\boldsymbol{P}^{\{\mathrm{A}\}}$) the rover's motion, accordingly to:

$$\begin{bmatrix} \boldsymbol{P}^{\{\mathrm{A}\}} \\ 1 \end{bmatrix} = {}^{\{A\}}_{\{B\}}\mathbf{T} \begin{bmatrix} \boldsymbol{P}^{\{\mathrm{B}\}} \\ 1 \end{bmatrix} = \begin{bmatrix} {}^{\{A\}}_{\{B\}}\mathbf{R} & \boldsymbol{\tau} \\ \mathbf{0}_{1\times 3} & 1 \end{bmatrix} \begin{bmatrix} \boldsymbol{P}^{\{\mathrm{B}\}} \\ 1 \end{bmatrix} \tag{2}$$

Hereafter, the left and right images of the first stereo pair (acquired before the motion step) will be denoted by $\mathcal{L}_1$ and $\mathcal{R}_1$, respectively. The symbols $\mathcal{L}_2$ and $\mathcal{R}_2$ will be used to refer to the images of the second stereo pair (acquired at the end of the drive step).

3. VISUAL ODOMETRY ALGORITHM

3.1 Feature Detection

A first step of VO algorithms is the identification of image keypoints that are matched and tracked across stereo pairs acquired at successive times. Image keypoints (*e.g.*, corners) are first detected in the stereo pair acquired at the beginning of the rover's motion step. The detection of such image-points should be robust to changes in the illumination conditions and the viewing angle of the scene. Corner-points are extracted using the Harris corner detector (Harris & Stephens, 1988), which identifies image pixels where the Harris score function gets a local maximum (*i.e.*, a pixel is classified as a corner if the associated Harris score is greater than the Harris scores computed at its 8 surrounding pixels).

To ensure a uniform distribution of corners across the image, it is divided in patches (or Region of Interest, ROI) that are processed independently, and the strongest corners in each ROI are selected (Figure 1). The usage of ROI also improves the corner detection in case of images including rover's structures, where corners associated with the metallic parts of the rover are much stronger than the ones associated with the environment (*i.e.*, rocks).

The extracted corner-points are associated with image-pixels and, therefore, have integer coordinates (x, y). Sub-pixels accuracies are attained through a least-squares fitting of a bivariate quadratic function $f(x, y)$ to the Harris metric responses computed in the (3×3) template window centered at the detected corner. The choice of a bivariate quadratic function is supported by the observation that the metric score about a corner has a distribution that can be locally fitted by a paraboloid (Zhu et al., 2007). The generic equation of the function $f(x, y)$ is:

$$f(x, y) = a_0x^2 + a_1y^2 + a_2x + a_3y + a_4xy + a_5 \tag{3}$$

where x and y are the column (sample) and row (line) coordinates of an image-point, respectively. In the least-squares fitting, the coordinates of the nine pixels inside the template window are remapped to be referred to the central pixel (x, y) (*e.g.*, the remapped coordinates of the central pixel are (0,0)). The least-squares estimate of the six polynomial coefficients $\hat{\boldsymbol{a}}$ is obtained accordingly to:

$$\hat{\boldsymbol{a}} = (\mathbf{B}^{\mathrm{T}}\mathbf{B})\mathbf{B}^{\mathrm{T}}\mathbf{F} \tag{4}$$

where $\mathbf{F}$ is the (9×1) column vector collecting the Harris score values associated with the nine image-points in the template window; and $\mathbf{B} = (\partial \boldsymbol{f}/\partial \boldsymbol{a})$ is the (9×6) matrix of the partial derivatives of the function f with respect to the polynomial coefficients $\boldsymbol{a} = [a_0, \dots, a_5]$ for the pixels in the template window. The remapping of the pixel coordinates yields a significant reduction of the computational cost related to $\mathbf{B}$. The fractional part of the refined corner-coordinates $(\Delta x, \Delta y)$ corresponds to the point where the function f is maximum, which is computed accordingly to:

$$\Delta x = -\frac{2a_1a_2 - a_3a_4}{4a_0a_1 - a_4^2} \tag{5}$$

$$\Delta y = -\frac{2a_0a_3 - a_2a_4}{4a_0a_1 - a_4^2} \tag{6}$$

The improved corner coordinates $(\hat{x}, \hat{y})$ are finally retrieved by adding the computed correction $(\Delta x, \Delta y)$ to the integer corner coordinates (x, y). Unreliable corners yielding corrections greater than 1 pixel are discarded.

A further down-selection is carried out to exclude corner-points at the image edges because of significant distortion effects. The landmarks associated with these corners are not well-suited to estimate the rover's pose since they may be off from the camera's field of view after the motion step. We discard corner-points that are within 30 pixels from the image boundaries.

3.2 Stereo-Matching

To enable the triangulation of the world-points, the extracted left and right corner-points are matched to find pairs of corners corresponding to the same landmark. To efficiently describe the neighborhood of the extracted corner-points, we adopt the SURF descriptor (Bay et al., 2006). The sum of squared differences (SSD) metric is used to compare the descriptors, and corner-points yielding the minimum SSD are matched. Since the images are not rectified, the epipolar constraint cannot be imposed, and the coordinates of left and right matched corners are expected to differ. However, this difference is assumed to be small, and to filter out wrong matches, pairs of matched corners with $|\hat{y}_L - \hat{y}_R| > 50$ pixels are discarded.

3.3 Triangulation

Each pair of stereo-matched corners is associated with a 3D world-point, whose coordinates can be retrieved by means of stereo triangulation. Our triangulation scheme accounts for the nonlinearities of the CAHVORE camera model (Gennery, 2006). The 3D coordinates of the world-points are determined as the midpoint of the minimum distance segment between two lines, which are the viewing rays projected out from the left and right entrance pupils. Under ideal conditions, the left and the right viewing rays exactly intersect at a point in space. In real cases they do not intersect because of image noise (that yields errors in the corner detection), matching errors, and camera model uncertainties. A minimum distance line segment connecting the two rays is detected, and the midpoint of the segment is taken as the best-estimated triangulated location of the landmark.

For rectified images described by the CAHV camera model (pinhole camera), the viewing rays depart from the projection center and intersect the image plane exactly where the corner-points are detected. To accurately retrieve the 3D landmarks coordinates in case of raw images, the nonlinear optical effects associated with the CAHVORE model are included. Given a pair of matched corners $\boldsymbol{p}_L$ and $\boldsymbol{p}_R$, their 2D coordinates are processed in combination with the CAHVORE parameters to adjust the locations of both left and right entrance pupils, $\boldsymbol{C}'_L$ and $\boldsymbol{C}'_R$, and viewing rays, $\boldsymbol{r}_L$ and $\boldsymbol{r}_R$ (Appendix A), which are unit vectors departing from $\boldsymbol{C}'_L$ and $\boldsymbol{C}'_R$, respectively (Gennery, 2006).

The 3D coordinates of the endpoints of minimum distance segment are defined as:

$$\begin{aligned} \boldsymbol{P}_L &= \boldsymbol{C}'_L + m_L \boldsymbol{r}_L \\ \boldsymbol{P}_R &= \boldsymbol{C}'_R + m_R \boldsymbol{r}_R \end{aligned} \tag{7}$$

where $m_L = \|\boldsymbol{P}_L - \boldsymbol{C}'_L\|$ and $m_R = \|\boldsymbol{P}_R - \boldsymbol{C}'_R\|$. The unknown parameters m_L and m_R are retrieved by enforcing that the minimum distance segment $(\boldsymbol{P}_R - \boldsymbol{P}_L)$ is orthogonal to the left and the right viewing unit vectors $\boldsymbol{r}_L$ and $\boldsymbol{r}_R$, as follows,

$$\begin{cases} (\boldsymbol{P}_R - \boldsymbol{P}_L) \cdot \boldsymbol{r}_L = 0 \\ (\boldsymbol{P}_R - \boldsymbol{P}_L) \cdot \boldsymbol{r}_R = 0 \end{cases}, \tag{8}$$

and, by substituting Eqs. (7) in Eqs. (8), we obtain:

$$\begin{cases} (\boldsymbol{C}'_R - \boldsymbol{C}'_L + m_R \boldsymbol{r}_R - m_L \boldsymbol{r}_L) \cdot \boldsymbol{r}_L = 0 \\ (\boldsymbol{C}'_R - \boldsymbol{C}'_L + m_R \boldsymbol{r}_R - m_L \boldsymbol{r}_L) \cdot \boldsymbol{r}_R = 0 \end{cases} \rightarrow \begin{cases} \boldsymbol{B} \cdot \boldsymbol{r}_L + m_R \boldsymbol{r}_R \cdot \boldsymbol{r}_L - m_L = 0 \\ \boldsymbol{B} \cdot \boldsymbol{r}_R + m_R - m_L \boldsymbol{r}_L \cdot \boldsymbol{r}_R = 0 \end{cases} \tag{9}$$

where $\boldsymbol{B} = \boldsymbol{C}'_R - \boldsymbol{C}'_L$ is the stereo baseline vector.

Eqs. (9) are solved for m_L and m_R providing the following solution:

$$\begin{cases} m_L = +\dfrac{\boldsymbol{B} \cdot \boldsymbol{r}_L - (\boldsymbol{B} \cdot \boldsymbol{r}_R)(\boldsymbol{r}_L \cdot \boldsymbol{r}_R)}{1 - (\boldsymbol{r}_L \cdot \boldsymbol{r}_R)^2} \\ \\ m_R = -\dfrac{\boldsymbol{B} \cdot \boldsymbol{r}_R - (\boldsymbol{B} \cdot \boldsymbol{r}_L)(\boldsymbol{r}_L \cdot \boldsymbol{r}_R)}{1 - (\boldsymbol{r}_L \cdot \boldsymbol{r}_R)^2} \end{cases} \tag{10}$$

The 3D coordinates of the landmark are then retrieved accordingly to:

$$\boldsymbol{P} = \frac{\boldsymbol{P}_L + \boldsymbol{P}_R}{2} \tag{11}$$

Since the 3D vectors $\boldsymbol{C}'_L$, $\boldsymbol{C}'_R$, $\boldsymbol{r}_L$ and $\boldsymbol{r}_R$ are defined with respect to the rover navigation frame, the 3D coordinates of the world-points $\boldsymbol{P}$ are referred to the rover navigation frame as well.

A parameter that measures the accuracy of the triangulated coordinates is the length d of the minimum distance segment, $d = \|\boldsymbol{P}_L - \boldsymbol{P}_R\|$. In our image processing algorithm, 3D point characterized by $d > 15$ cm are filtered out as outliers (Figure 2).
The (3×3) covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{P}}$ associated with the triangulated point is retrieved by propagating the (2×2) covariances related to the left and right corner-points, $\boldsymbol{\Sigma}_{\boldsymbol{p}_L}$ and $\boldsymbol{\Sigma}_{\boldsymbol{p}_R}$. In this work, we assume that $\boldsymbol{\Sigma}_{\boldsymbol{p}_L} = \boldsymbol{\Sigma}_{\boldsymbol{p}_R} = \sigma^2 \mathbb{I}_{2\times 2}$, with $\sigma = 0.5$ pixels. $\boldsymbol{\Sigma}_{\boldsymbol{P}}$ is then computed accordingly to:

$$\boldsymbol{\Sigma}_{\boldsymbol{P}} = \mathbf{J}\boldsymbol{\Sigma}_{\boldsymbol{p}}\mathbf{J}^{\mathrm{T}}, \qquad \boldsymbol{\Sigma}_{\boldsymbol{p}} = \begin{bmatrix} \boldsymbol{\Sigma}_{\boldsymbol{p}_L} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_{\boldsymbol{p}_R} \end{bmatrix} \tag{12}$$

where $\mathbf{J}$ is the (3×4) Jacobian matrix defined as:

$$\mathbf{J} = \begin{bmatrix} \dfrac{\partial \boldsymbol{P}}{\partial x_L} & \dfrac{\partial \boldsymbol{P}}{\partial y_L} & \dfrac{\partial \boldsymbol{P}}{\partial x_R} & \dfrac{\partial \boldsymbol{P}}{\partial x_R} \end{bmatrix} \tag{13}$$

We computed the columns of the Jacobian matrix by recursively applying the chain rule for the partial derivatives. We provide the analytical expressions of the derived Jacobian matrix $\mathbf{J}$ in Appendix A. Hereafter, the symbols $\boldsymbol{P}^{\{\mathrm{B}\}}$ and $\boldsymbol{\Sigma}_{\boldsymbol{P}}^{\{\mathrm{B}\}}$ will be used to denote the landmarks coordinates triangulated before the motion step and the associated covariance matrix, respectively. The corresponding quantities computed after the motion step will be denoted by $\boldsymbol{P}^{\{\mathrm{A}\}}$ and $\boldsymbol{\Sigma}_{\boldsymbol{P}}^{\{\mathrm{A}\}}$.
The (3×3) covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{P}}$ is defined with respect to the rover's frame, and reflects the 3D distribution of the uncertainties related to the triangulated world-point coordinates. As a first approximation, the 3D point covariance is assumed Gaussian, and can be represented as an ellipsoid elongated along the line-of-sight direction from the camera to the landmark. Figure 3 shows the 1-σ formal uncertainties associated with the X-, Y-, Z-coordinates of the retrieved 3D points expressed in the left NavCam frame {L} (that is almost aligned with the right NavCam frame {R}). The camera frame {L} is defined as follows: +Z-axis along the camera optical axis, pointing outwards; +Y-axis along the image central column and pointing towards the top row; +X-axis along the image central row and pointing towards the image left column. To be consistent with the selected frame, the 3D points covariances $\boldsymbol{\Sigma}_{\boldsymbol{P}}$ are transformed accordingly to:

$${}^{\{\mathrm{L}\}}\boldsymbol{\Sigma}_{\boldsymbol{P}} = {}^{\{\mathrm{L}\}}_{\{\mathrm{N}\}}\mathbf{R}\ \boldsymbol{\Sigma}_{\boldsymbol{P}}\ \left({}^{\{\mathrm{L}\}}_{\{\mathrm{N}\}}\mathbf{R}\right)^{\mathrm{T}} \tag{14}$$

where ${}^{\{\mathrm{L}\}}_{\{\mathrm{N}\}}\mathbf{R}$ is the rotation matrix from the rover navigation frame {N} to the left NavCam frame {L}, which is obtained from the attitude mission kernels. The 1-σ formal uncertainties are then retrieved by taking the square root of the elements along ${}^{\{\mathrm{L}\}}\boldsymbol{\Sigma}_{\boldsymbol{P}}$ principal diagonal. As expected, a strong correlation of the uncertainties with the relative distance of the landmarks from the rover is observed (*i.e.*, the farther the landmarks are, the greater the uncertainties are). The main contribution is related to σ_z (Figure 3c), since the camera boresight (*i.e.*, Z-axis) is mainly aligned with the line-of-sight direction. Compared to σ_z, the uncertainties on the X- and Y-coordinates show a greater dependence on the relative orientation of the line-of-sight and the image horizontal (*i.e.*, X-axis) and vertical (*i.e.*, Y-axis) directions, although dominant variations are associated with the distance of the landmarks from the rover. σ_x (Figure 3a) and σ_y (Figure 3b) are observed to increase towards the lateral boundaries and the top of the image, respectively.

3.4 Tracking

To predict the 3D coordinates of the triangulated landmarks after the motion step, we would need to directly propagate the rover's pose. Since WO and IMU measurements are not archived, we are not able to directly accomplish this task. However, we update the landmarks 3D coordinates after the triangulation by using the pose information that are included in the image metadata. These preliminary estimates were retrieved onboard the vehicle by processing WO and IMU data. The rover's position and orientation are reported with respect to the site frame {S}, which is a fixed coordinate frame attached to the Martian surface. The center of this frame is periodically updated by the surface operations team to mitigate accumulation of the rover position errors. Ancillary information regarding the pose of the rover's navigation frame with respect to the site frame are included in the image metadata as position vector $\boldsymbol{P}^{\{\mathrm{S}\}}$ and attitude quaternion $\boldsymbol{q}$. The motion parameters associated with the rover's motion are retrieved from the telemetered rover's pose according to:

$$
\begin{gathered}
{}^{\{\mathrm{A}\}}_{\{\mathrm{B}\}}\mathbf{R} = {}^{\{\mathrm{A}\}}_{\{\mathrm{S}\}}\mathbf{R}\left({}^{\{\mathrm{B}\}}_{\{\mathrm{S}\}}\mathbf{R}\right)^{\mathrm{T}} \\
\boldsymbol{\tau} = {}^{\{\mathrm{A}\}}_{\{\mathrm{S}\}}\mathbf{R}\left(\boldsymbol{P}^{\{\mathrm{S}\}}_{\mathrm{B}} - \boldsymbol{P}^{\{\mathrm{S}\}}_{\mathrm{A}}\right)
\end{gathered} \tag{15}
$$

where $\left(\boldsymbol{P}^{\{\mathrm{S}\}}_{\mathrm{B}}, {}^{\{\mathrm{B}\}}_{\{\mathrm{S}\}}\mathbf{R}\right)$ and $\left(\boldsymbol{P}^{\{\mathrm{S}\}}_{\mathrm{A}}, {}^{\{\mathrm{A}\}}_{\{\mathrm{S}\}}\mathbf{R}\right)$ are the position vector and the rotation matrix defining the rover's pose (with respect to the site frame) before and after the drive step, respectively. Matrices ${}^{\{\mathrm{B}\}}_{\{\mathrm{S}\}}\mathbf{R}$ and ${}^{\{\mathrm{A}\}}_{\{\mathrm{S}\}}\mathbf{R}$ are retrieved from the associated quaternions.

The tracking step identifies in the new left image $\mathcal{L}_2$ (acquired at the end of the drive step) the corner-points associated with the landmarks observed before the rover's motion. To accomplish this task, the updated 3D points are first projected onto $\mathcal{L}_2$ accounting for the nonlinearities of the CAHVORE camera model, yielding a 2D point $\bar{\boldsymbol{p}}^{i}_{\mathcal{L}_2}$ for each feature $i = 1, \dots, N_B$, with N_B that denotes the number of triangulated landmarks before the motion step. A local corner detection (within a 21×21 pixels search region) is carried out about each point $\bar{\boldsymbol{p}}^{i}_{\mathcal{L}_2}$ to extract the keypoints $\boldsymbol{p}^{i,k}_{\mathcal{L}_2}$ $(k = 1, \dots, N_i)$ that can be associated with $\boldsymbol{\mathcal{P}}_i$. To enable an accurate match of keypoints between the left images before and after the motion step, we adopted a Normalized Cross Correlation (NCC)-based strategy, and square template windows of the same size $\mathcal{W}^{i}_{\mathcal{L}_1}$ and $\mathcal{W}^{i,k}_{\mathcal{L}_2}$ are defined about $\boldsymbol{p}^{i}_{\mathcal{L}_1}$ (*i.e.*, the corner-point associated with $\boldsymbol{\mathcal{P}}_i$ and detected in the first left image) and each of the locally detected corners $\boldsymbol{p}^{i,k}_{\mathcal{L}_2}$ in the second left image, respectively. The template window $\mathcal{W}^{i}_{\mathcal{L}_1}$ is then compared to each of the N_i template windows $\mathcal{W}^{i,k}_{\mathcal{L}_2}$ $(k = 1, \dots, N_i)$ accordingly to the NCC index, defined as:

$$
\mathrm{NCC}_k = \frac{\sum_{j=1}^{n}\left[\mathrm{DN}_1(j)\right]\left[\mathrm{DN}_2^k(j)\right]}{\sqrt{\sum_{j=1}^{n}\left[\mathrm{DN}_1(j)\right]^2 \sum_{j=1}^{n}\left[\mathrm{DN}_2^k(j)\right]^2}} \tag{16}
$$

where n is the number of the pixels in a single template window; and $\mathrm{DN}_1(j)$ and $\mathrm{DN}_2^k(j)$ are the Digital Numbers (DN) associated with the j^{th} pixel in $\mathcal{W}^{i}_{\mathcal{L}_1}$ and $\mathcal{W}^{i,k}_{\mathcal{L}_2}$, respectively. The locally detected corner-point that yields the maximum NCC is assumed to be the keypoint (in $\mathcal{L}_2$) associated with $\boldsymbol{\mathcal{P}}_i$. In this work, we used 11×11 template windows, imposing a minimum NCC threshold of 0.85 to discard unreliable tracked corners.

3.5 3D-to-3D Motion Estimate

A second stereo matching is carried out to match the corners tracked in the second left image $\mathcal{L}_2$ with the corner-points extracted from the second right image $\mathcal{R}_2$. The 3D coordinates of the associated landmarks ($\boldsymbol{P}_i^{\{A\}}$) are triangulated after the motion, and their covariances ($\boldsymbol{\Sigma}_i^{\{A\}}$) are computed.

At the end of this step, two 3D point-clouds are obtained; they are made up of the same set of landmarks observed before and after the rover's motion. The 3D-to-3D VO algorithm processes the two sets of 3D points, providing a maximum-likelihood estimate of the rover's rototranslation that best-aligns the point-clouds.

An initial estimate of the rover's motion (${}^{\{A\}}_{\{B\}}\widehat{\mathbf{R}}_0, \hat{\boldsymbol{\tau}}_0$) is obtained through a least-squares solution (Arun et al., 1987). The first-guess solution is then refined through the maximum-likelihood estimation (MLE) algorithm, which minimizes the cost function U depending on the residuals $\boldsymbol{e}_i = \boldsymbol{P}_i^{\{A\}} - {}^{\{A\}}_{\{B\}}\mathbf{R}\, \boldsymbol{P}_i^{\{B\}} - \boldsymbol{\tau}$:

$$U = \sum_{i=1}^{N_{LM}} \left(\boldsymbol{e}_i^{\mathrm{T}}\, \mathbf{W}_i\, \boldsymbol{e}_i \right) \tag{17}$$

with N_{LM} denoting the number of landmarks that are identified before and after the drive step. The residuals are weighted using the (3×3) matrix $\mathbf{W}_i$, which accounts for the covariance matrices associated with the triangulated points $\boldsymbol{\Sigma}_i^{\{B\}}$ and $\boldsymbol{\Sigma}_i^{\{A\}}$ (Matthies & Shafer, 1987), accordingly to:

$$\mathbf{W}_i = \left(\boldsymbol{\Sigma}_i^{\{A\}} + {}^{\{A\}}_{\{B\}}\mathbf{R}\, \boldsymbol{\Sigma}_i^{\{B\}} \left({}^{\{A\}}_{\{B\}}\mathbf{R} \right)^{\mathrm{T}} \right)^{-1} \tag{18}$$

The inverse of $\mathbf{W}_i$ is the covariance matrix associated with the residual errors obtained by linearizing the motion equation (Eq. (2)) about a first-guess solution for the rover's motion. The MLE is iterated until convergence that is declared when the quantity $|\boldsymbol{\Theta}_t - \boldsymbol{\Theta}_{t-1}|$ is lower than a tolerance of 10^{-6} radians, with t denoting the current iteration. The first point-cloud $\boldsymbol{P}_i^{\{B\}}$ $(i = 1, \ldots, N_{LM})$ is then transformed accordingly to the retrieved maximum-likelihood solution $\left({}^{\{A\}}_{\{B\}}\widehat{\mathbf{R}}, \hat{\boldsymbol{\tau}} \right)$, and the resulting 3D points are projected back onto the second stereo pair, enabling the computation of the reprojection error. The updated coordinates of the i^{th} landmark are defined as:

$$\overline{\boldsymbol{P}}_i^{\{A\}} = {}^{\{A\}}_{\{B\}}\widehat{\mathbf{R}} \boldsymbol{P}_i^{\{B\}} + \hat{\boldsymbol{\tau}}, \tag{19}$$

and the reprojection error is computed accordingly to:

$$E_i = \left| \boldsymbol{p}_{\mathcal{L}_2}^i - \overline{\boldsymbol{p}}_{\mathcal{L}_2}^i \right| + \left| \boldsymbol{p}_{\mathcal{R}_2}^i - \overline{\boldsymbol{p}}_{\mathcal{R}_2}^i \right| \tag{20}$$

where $\overline{\boldsymbol{p}}_{\mathcal{L}_2}^i$ and $\overline{\boldsymbol{p}}_{\mathcal{R}_2}^i$ are the 2D points retrieved by reprojecting $\overline{\boldsymbol{P}}_i^{\{A\}}$ onto $\mathcal{L}_2$ and $\mathcal{R}_2$, respectively. Landmarks that show $E_i > E^{MAX}$ are filtered out as outliers, since they are based on mismatched or mistracked corner-points. The down-selected landmarks are then used to compute the next maximum-likelihood solution. The value of the threshold E^{MAX} is fixed to 30 pixels for the first iteration and is reduced by $\Delta E = 5$ pixels at each iteration. The estimation procedure is iterated until the reprojection error is lower than 5 pixels for each keypoint, leading to a maximum of six iterations. This tuning scheme was implemented to minimize the computational time required to declare convergence of the algorithm and to enable high accuracies of the rover's pose. A looser threshold E^{MAX} is initially adopted to discard outstanding outliers that may lead to the solution divergence. By reducing the projection tolerance by ΔE at each iteration, a refined exclusion of the remaining outliers is obtained to enhance the pose reconstruction.

After processing a set of stereo images captured during a traverse, the trajectory of the rover is retrieved by sequentially linking the estimated motion steps. The rover's position with respect to the site frame $\{\mathrm{S}\}$ at epoch t_k (*i.e.*, at the end of the k^{th} drive step) is defined as:

$$\boldsymbol{P}_k^{\{\mathrm{S}\}} = \boldsymbol{P}_{k-1}^{\{\mathrm{S}\}} + {}^{\{\mathrm{S}\}}_{\{\mathrm{k-1}\}}\mathbf{R}\left({}^{\{\mathrm{k}\}}_{\{\mathrm{k-1}\}}\mathbf{R}\right)^{\mathrm{T}}\left(-\boldsymbol{\tau}^{\{\mathrm{k}\}}\right), \tag{21}$$

where $\boldsymbol{P}_{k-1}^{\{\mathrm{S}\}}$ is the rover's location at epoch t_{k-1} (*i.e.*, before the k^{th} drive step); ${}^{\{\mathrm{S}\}}_{\{\mathrm{k-1}\}}\mathbf{R}$ is the rotation matrix from the rover navigation frame at epoch t_{k-1} to the site frame $\{\mathrm{S}\}$; ${}^{\{\mathrm{k}\}}_{\{\mathrm{k-1}\}}\mathbf{R}$ is the maximum-likelihood estimated rotation matrix between the rover navigation frame before ({k-1}) and after ({k}) the k^{th} motion step; and $\boldsymbol{\tau}^{\{\mathrm{k}\}}$ is the MLE-estimated translation vector associated with the k^{th} motion step. The updated rover's orientation with respect to the site frame $\{\mathrm{S}\}$ is retrieved as:

$${}^{\{\mathrm{k}\}}_{\{\mathrm{S}\}}\mathbf{R} = {}^{\{\mathrm{k}\}}_{\{\mathrm{k-1}\}}\mathbf{R}\left({}^{\{\mathrm{S}\}}_{\{\mathrm{k-1}\}}\mathbf{R}\right)^{\mathrm{T}}. \tag{22}$$

The MLE provides the (6×6) covariance matrix associated with the rover's position and orientation variations during the motion step ($\boldsymbol{\Sigma}_{k-1}^{k}$). To propagate the formal uncertainties of the rover's pose (Figure 4), this matrix is combined with the covariance matrix obtained after the previous motion step ($\boldsymbol{\Sigma}_{k-1}$), as follows:

$$\boldsymbol{\Sigma}_k = \mathbf{J}_{C,k}\begin{bmatrix}\boldsymbol{\Sigma}_{k-1} & \mathbf{0}_{6\times 6}\\ \mathbf{0}_{6\times 6} & \boldsymbol{\Sigma}_{k-1}^{k}\end{bmatrix}\mathbf{J}_{C,k}^{\mathrm{T}}, \tag{23}$$

where $\mathbf{J}_{C,k}$ is the Jacobian matrix,

$$\mathbf{J}_{C,k} = \begin{bmatrix}\dfrac{\partial \boldsymbol{C}_k}{\partial \boldsymbol{C}_{k-1}} & \dfrac{\partial \boldsymbol{C}_k}{\partial \boldsymbol{M}_{k-1}^{k}}\end{bmatrix}. \tag{24}$$

The partial derivatives in the Jacobian matrix are computed between the pose vector $\boldsymbol{C}_k$ and the pose vector of the previous step, $\boldsymbol{C}_{k-1}$, and the estimated motion parameters vector, $\boldsymbol{M}_{k-1}^{k} = [\widehat{\boldsymbol{\Theta}} \quad \widehat{\boldsymbol{\tau}}]$, which defines the k^{th} rototranslational motion step.

The estimation of the rover's pose uncertainty is a highly nonlinear problem. The mathematical formulation adopted in this study may then be affected by errors associated with the linearization of the equations used to update the rover's pose. Furthermore, it relies on the strong assumption that the uncertainty of the rover's pose is Gaussian, but the true probability distribution of the rover's state vector may be non-Gaussian. These factors deeply affect the evolution of the pose covariance over the sequence of motion steps, and the accumulation of errors will eventually produce inconsistent results, *i.e.*, the estimated uncertainty is smaller than the true error (Bailey et al., 2006). Therefore, the results retrieved by standard linearization-based methods only hold in first approximation, and for an accurate uncertainty propagation over longer motion sequences Monte Carlo-based approaches should be adopted instead (Pertile et al., 2014). However, due to the limited number of concatenated motion steps for each traverse, the standard linearized formulation is still suitable and adopted in the study.

4. TEST AND VALIDATION OF THE VO ALGORITHM: MER IMAGES ANALYSIS

The 3D-to-3D VO algorithm was used to process two pairs of stereo images acquired by the NavCams of the NASA MER-1 rover Opportunity (Maki et al., 2003) on sol 839 (4 June 2006) and 840 (5 June 2006). These optical data are 8-bit full-resolution (*i.e.*, 1024×1024 pixel) images. To be consistent with MER VO flight software, we analyzed rectified images derived from raw data, which are described by a simplified CAHV camera model (on the NASA Planetary Data System (PDS) archive, the *rectified* images are named *linearized* products). In the CAHV model, the vector $\boldsymbol{O}$ is not defined as it coincides with $\boldsymbol{A}$, and the three vectors $\boldsymbol{A}$, $\boldsymbol{H'}$ and $\boldsymbol{V'}$ are mutually orthogonal and form a right-handed frame attached to the image plane. When the pinhole camera model is adopted, closed-form equations can be used to triangulate the world-points (*e.g.*, Matthies & Shafer, 1987; Andolfo et al., 2021; Andolfo et al. 2022).

A first reprojection of the 3D point-cloud retrieved before the motion step onto the second stereo pair is carried out with the rover's trajectory and attitude archived in the mission SPICE kernels. Figure 5b shows the set of these 2D points (red dots) that are not consistent with the landmarks observed in the first stereo pair (blue dots in Figure 5a). These discrepancies preserve crucial information on the errors of the rover's telemetered position, and are then used in our MLE motion estimate to enhance the knowledge of the rover's drive step. Our VO algorithm allowed us to adjust the rover's position, leading to an updated reprojection of the point cloud (green dots in Figure 5b) that is in full agreement with the landmarks observed before the motion. The reconstructed motion indicates that the length of the path travelled by the rover differs by ~7 cm compared to the WO-based estimate. No significant corrections are observed for the rotation vector that is in line with the orientation provided by the IMU (~0.1° error).

To cross check our results, we also analyzed the point-clouds archived on the NASA PDS as XYZ images associated with the selected stereo pairs. These 3-bands images provide (X, Y, Z) Cartesian coordinates of the world-points in the field of view of the rover's left NavCam (Chen, 2014). The 3D coordinates of the surface points are defined with respect to the site frame. Since objects, such as rocks and boulders, are fixed on the surface, their coordinates with respect to the site frame do not change before and after the motion step. The archived point-clouds, however, show different 3D coordinates of a set of world-points (Figure S1) retrieved before and after the motion step (Supplementary Tables 1−2). The discrepancies between the archived point clouds before and after the motion, which were caused by dead-reckoning errors, are fully consistent with our motion adjustment based on the more accurate VO-based method.

5. RESULTS

A set of raw non-rectified stereo images acquired by Perseverance NavCams were processed with our 3D-to-3D VO algorithm. The selected images were acquired on sols 65 (26 April 2021), 66 (27 April 2021) and 72 (3 May 2021), and are single-tile 1280×960 pixels images (*i.e.*, 4× downsampling). On these sols, the rover drove across the Van Zyl overlook region, a favorable surface area for the monitoring of the first Ingenuity helicopter's flight attempts, and for testing its AutoNav driving capabilities. To limit pose reconstruction issues, we partitioned the rover's trajectory into different legs that exclude strong changes in the heading angle. The lack of data during these surface maneuvers would prevent us from precise recovery of the rover's pose.

We also analyzed the stereo pairs acquired on sol 120 (22 June 2021), when the rover took images of the surrounding environment changing the cameras pointing direction only. These images are characterized by a 2× downsampling, and the full images are made up of four tiles with a size of 1280×960 pixels. To process them, we preliminary assembled the tiles, obtaining 2560×1280 pixels images. A complete list of the processed images is reported in the Supplementary Table 3.

5.1 Traverse Cases: sol 65, 66, 72

The 3D-to-3D VO algorithm was used to process sequences of stereo images acquired along the path. By combining the estimated drive steps accordingly to Eqs. $(21) - (22)$, we retrieved an interpolated continuous trajectory for each leg.

The rover's path reconstructed by using our VO software is shown in Figure 6, and the points displayed along the trajectories show the estimated locations where new stereo pairs were acquired. At the beginning of sol 65, Perseverance is located at the North-West area of the Van Zyl overlook region. Perseverance was then commanded to move Eastwards (sol 65, red), and during this first driving session the rover maintained the cameras pointed in the opposite direction with respect its motion. To adjust its heading direction, once completing the first traverse, the rover performed a ~90° turn-in-place rotation about the yaw axis, and then started driving towards the South-East area (sol 65, green). The route was maintained in the next sols (sol 66, blue; sol 72, black), while slight adjustments to the heading direction were carried out during the motion. The major number of stereo images was taken on sol 72, when the rover travelled the longest distance (>11.5 meters).

To quantify the difference between our VO solution and the telemetry-based rover's path, we computed the distance between the telemetered and the estimated rover's location at each new stereo pair acquisition. We retrieved Perseverance's telemetry data from two main sources, *i.e.*, image metadata, and telemetry data archived in the Position Localization and Attitude Correction Estimate Storage (PLACES) database (Deen, 2022). Perseverance can use several localization techniques (*i.e.*, wheel rotation integration, sun find) to update its position and attitude, and the pose solutions produced onboard (*i.e.*, not adjusted by ground operators) are all tracked in the PLACES database. After completing a motion step, if the rover refines its location or attitude by using, for example, VO techniques, the pose counter of Rover Motion Counter (RMC) is incremented, and the new localization solution is stored in the PLACES database. In addition to the position and attitude data reported in the image labels (*i.e.*, first localization solution), refined pose estimations are available for sol 72, and we considered them to further compare the solutions.

As shown in Figure 7, position discrepancies accumulate on average as the rover moves for all the estimated traverses. This trend can be observed by using the location and attitude data reported in the image labels (solid), and the refined pose estimates retrieved from the PLACES database (dashed). Major increments of position discrepancies (>20 cm) are observed when adjacent stereo pairs are acquired more than 1-meter apart.

For sol 72, we observe that our VO-solution is more consistent with the refined pose estimated by the rover (black, dashed) compared to the solution based on WO and IMU data only (black, solid). Furthermore, we computed the difference between the estimated distance travelled at each drive step based on our 3D-to-3D VO solution and the refined telemetry data (Figure 8). We observe that short distances (<1 m) traversed by the rover lead to a full agreement between the two independent solutions (*i.e.*, differences of 1-2 cm). Larger discrepancies (~20 cm) are detected for cumulative drive steps >3.5 meters. The estimation of such large drive steps through computer vision techniques is extremely challenging, as adjacent images should be sufficiently overlapped. For the MER rovers, for example, turn-in-place rotations and forward steps were limited to 18° and to 75 cm, respectively (Matthies et al., 2007). Larger surface maneuvers result in significant differences between looking angles and image resolution that may affect the tracking of keypoints across stereo pairs, inducing possible larger errors in the pose reconstruction.

The discrepancies may also rely on the weights adopted in the MLE estimation. By scaling the landmarks covariances accordingly to the landmarks distance from the rover, modified weighting matrices can be computed as:

$$
\begin{gathered}
\overline{\boldsymbol{\Sigma}}_i^{\{\mathrm{B}\}} = \left|\boldsymbol{P}_i^{\{\mathrm{B}\}}\right|^2 \boldsymbol{\Sigma}_i^{\{\mathrm{B}\}} \qquad \overline{\boldsymbol{\Sigma}}_i^{\{\mathrm{A}\}} = \left|\boldsymbol{P}_i^{\{\mathrm{A}\}}\right|^2 \boldsymbol{\Sigma}_i^{\{\mathrm{A}\}} \\
\overline{\mathbf{W}}_i = \left(\overline{\boldsymbol{\Sigma}}_i^{\{\mathrm{A}\}} + {}_{\{\mathrm{B}\}}^{\{\mathrm{A}\}}\mathbf{R}\,\overline{\boldsymbol{\Sigma}}_i^{\{\mathrm{B}\}}\left({}_{\{\mathrm{B}\}}^{\{\mathrm{A}\}}\mathbf{R}\right)^{\mathrm{T}}\right)^{-1}
\end{gathered} \quad . \tag{25}
$$

This different weighting of the measurements leads to position estimates that are more consistent with the telemetered trajectory for motion steps >1 m. By including the landmark distance in the data weighting, the MLE adjustment of the rover's position is limited if the pairs of stereo images are acquired at relative distant locations, since the tracked features are far apart from the rover. The 3D-point measurements are then deweighted for longer motion steps.

To further investigate the relationship between the position discrepancies and the length of the drive steps, we employed the 3D-to-3D VO scheme to estimate some extended motion steps that we defined by combining multiple motion steps into a single longer drive step. Extended drive steps were retrieved on sol 66 (*i.e.*, by combining the drive steps 3-4) and on sol 72 (*i.e.*, by combining the drive steps 3-4-5, and the drive steps 7-8-9, separately), and all have length of ~2 meters. We then compared the motion estimates produced by the VO algorithm by considering the single extended motion steps, and the multiple shorter drive steps.

For sol 72, the longer drive steps yield position discrepancies that are comparable to the ones obtained by considering multiple shorter drive steps separately, while for sol 66, the extended drive step produces significantly larger position differences (~35 cm) compared to the case in which the drive steps are considered separately (~15 cm). These discrepancies result from the different tracking process of the terrain features. The combination of surface morphology and illumination conditions significantly affects the detection of the surface landmarks that are then tracked along the path. The traverses during sol 72 that are analyzed by excluding intermediate stereo-pairs are characterized by higher pose estimation accuracies since the algorithm tracks successfully features that are well-defined by the solar illumination on the bedrock terrain at the local time of the image acquisition (Figure S2). The associated keypoints are well-defined corners outlined by the contrast between the bright small rocks and the shadows they cast on the terrain. These robust 3D-3D correspondences are processed by the MLE filter, which yields motion estimates consistent with the case in which small drive steps are considered separately, and then combined. On the other hand, for the traverse of sol 66, the terrain close to the rover is quite repetitive and hosts smooth rocks, leading to some erroneous 3D-3D correspondences that affect the reconstructed motion, and produce greater discrepancies.

The results suggest that if there are some peculiar landmarks in the scene to which anchor robust 3D-3D correspondences, the estimation of longer (~2 m) drive steps may not inflate significant errors in the reconstructed path. However, smaller drive steps are preferred, since it is easier for the algorithm to track more keypoints that can be processed in the motion estimation filter. Further analyses will be carried out as new suitable sequences of stereo NavCam pairs are archived on the NASA PDS.

To assess the level of accuracy of our 3D-to-3D VO estimated trajectory, we employed a loop closing technique based on the matching of common features that are observed in the first- and the last-acquired stereo pair along each traverse. We first carried out separate stereo-matching of detected corner-points in the first and the last stereo pairs. Image-points associated with the same landmarks were then identified in the two stereo pairs, and 3D coordinates of the selected landmarks were triangulated at the beginning ($\boldsymbol{P}_i^{\{\mathrm{b}\}}$) and at the end ($\boldsymbol{P}_i^{\{\mathrm{e}\}}$) of the overall rover's traverse. The telemetered and reconstructed position and attitude of the rover were used to retrieve the 3D points $\widehat{\boldsymbol{P}}_i^{\{\mathrm{e}\}}$, which are computed through the transformation of $\boldsymbol{P}_i^{\{\mathrm{b}\}}$ according to the telemetered and estimated $\mathbf{R}$ and $\boldsymbol{\tau}$, as follows:

$$\widehat{\boldsymbol{P}}_i^{\{e\}} = \mathbf{R}\boldsymbol{P}_i^{\{b\}} + \boldsymbol{\tau}. \tag{26}$$

By subtracting $\widehat{\boldsymbol{P}}_i^{\{e\}}$ to $\boldsymbol{P}_i^{\{e\}}$, we obtain the distance $D_i = \left|\boldsymbol{P}_i^{\{e\}} - \widehat{\boldsymbol{P}}_i^{\{e\}}\right|$ between the triangulated and the estimated 3D points location after the motion. The lower this parameter is, the more accurate the retrieved position and attitude of the rover are. By selecting a set of landmarks, we computed the parameter D_i for each landmark with the telemetered trajectory (D^{TLM}) and the trajectory retrieved through our VO algorithm (D^{VO}). The results indicate that all paths reconstructed through our VO software provide smaller errors compared to the telemetry-based trajectories (*i.e.*, $D^{TLM}/D^{VO} > 1$), supporting a more accurate reconstruction of the rover's path through our localization solution (Table 1). Landmarks position discrepancies computed accordingly to our VO-reconstructed path are reduced by a factor >2 compared to the telemetry-based solution. The left images acquired at the beginning and at the end of the estimated paths are shown in Figure 9 and Figures S3−S6.

Table 1. Landmarks position discrepancies computed accordingly to our VO-reconstructed path (D^{VO}) and the telemetered rover's path (D^{TLM}).

	Landmark ID	1	2	3	4
Sol 65 (1st leg)	$\boldsymbol{D^{VO}}$ [m]	0.018	0.054	0.017	0.023
	$\boldsymbol{D^{TLM}}$ [m]	0.065	0.118	0.068	0.074
	$\boldsymbol{D^{TLM}/D^{VO}}$	**3.69**	**2.19**	**4.02**	**3.20**
Sol 65 (2nd leg)	$\boldsymbol{D^{VO}}$ [m]	0.073	0.034	0.092	0.101
	$\boldsymbol{D^{TLM}}$ [m]	0.292	0.332	0.455	0.275
	$\boldsymbol{D^{TLM}/D^{VO}}$	**4.01**	**9.81**	**4.94**	**2.73**
Sol 66	$\boldsymbol{D^{VO}}$ [m]	0.189	0.071	0.214	0.075
	$\boldsymbol{D^{TLM}}$ [m]	0.878	0.779	0.941	0.711
	$\boldsymbol{D^{TLM}/D^{VO}}$	**4.65**	**10.91**	**4.40**	**20.57**
Sol 72	$\boldsymbol{D^{VO}}$ [m]	0.158	0.155	0.216	0.167
	$\boldsymbol{D^{TLM}}$ [m]	0.431	0.604	0.712	0.348
	$\boldsymbol{D^{TLM}/D^{VO}}$	**2.74**	**3.90**	**3.29**	**2.08**

5.2 Rotation case: sol 120

On sol 120, six stereo pairs were acquired by the rover from the same spot by only changing the NavCams pointing direction through the azimuth and elevation angles of the rover's mast. This case is an important test to better understand whether our algorithm may deal with pure rotations. As the position and the attitude of the rover navigation frame with respect to the site frame were kept fixed, the rotation and the translation vectors to be estimated are both null. The results show that the attitude error is always <0.1° along the yaw (blue), pitch (green) and roll (red) axes (Figure 10a), which is fully in line with the expected motion, and the overall position discrepancy (black) is <1.5 cm (Figure 10b). The small position discrepancies can be explained by a few erroneous 3D-3D correspondences, which impact on the final solution, and by the estimation filter itself, which is not constrained to provide a null translation.

5.3 Discussion

The past and current Mars surface exploration missions have demonstrated a successful use of vision-based localization techniques to provide reliable and accurate pose estimates that are required to optimally plan the rover's activities. VO methods enable to estimate the rover's change-in-pose by only processing the camera measurements that are independent from dead-reckoning data, and are not affected by the wheel-soil interaction forces. However, the image processing steps are time-consuming, and a dedicated hardware (*e.g.*, Field Programmable Gate Array, FPGA) is required to support real-time applications, as on Perseverance (Verma et al., 2022).

Although VO methods can mitigate localization errors related to the wheels slippage, they are still dependent on the terrain characteristics, and good visual inputs are required to produce reliable motion estimates. For example, untextured or repetitive terrains are bad for the keypoints extraction and tracking, and may yield to the algorithm convergence failure. Also, optical cameras are not suitable to navigate at night. On the other hand, dead-reckoning methods based on wheel encoders and IMU data are independent from the illumination conditions, although visual obstacle detection is disabled, jeopardizing the rover's safety.

By providing relative pose updates only, VO techniques are not suitable to recover the rover's path over very long traverses, since localization errors will eventually accumulate, representing a risk for safe rover operations. To reduce the drift of the reconstructed trajectory from the real path, additional measurements provided by other onboard instruments can be included in the localization algorithm. Sun sensor (Olson et al., 2003) and star-trackers (Enright et al., 2012) data, for example, can provide periodical updates to the absolute rover's orientation, limiting the growth rate of position errors. Auxiliary orientation data can also be beneficial to adjust the rover's attitude after large reorientation maneuvers, as the one performed by Perseverance on sol 65. To further mitigate the accumulated errors during the motion, a joint refinement of the rover's positions and the landmarks coordinates (*i.e.*, keyframe-based bundle adjustment) is usually carried out at the end of the rover's driving sessions. By minimizing the reprojection error associated with the observed landmarks in multiple images, these methods allow to refine the 3D locations of the observed features, and to significantly increase the consistency of the reconstructed path, enabling high localization accuracies through extended traverses (Di et al., 2008).

Differently from terrestrial applications, planetary rovers typically carry out one-way traverses (Matthies et al., 2007), due to the main requirement of exploring new areas of scientific interest, and accounting for the limited rover's speed. In the next years, novel highly movable rovers (speed >20 cm/s) will be launched, such as the lunar NASA VIPER rover (Utz & Fluckiger, 2021), and the advanced moving capabilities of these assets will enable the rovers to revisit the same areas multiple times. Simultaneous Localization and Mapping (SLAM) approaches could then take advantage of loop-closure detection to globally adjust the rover's trajectory while building a consistent map of the operational environment (Hidalgo-Carrió et al., 2018; Giubilato et al., 2022), paving the way for an effective human-robotic cooperative framework.

6. CONCLUSIONS

Trajectory reconstructions of the NASA Mars 2020 rover Perseverance were presented in this study to enable a better understanding of the vehicle's path, when the rover explored the Van Zyl overlook region. After introducing the CAHVORE camera parameters, which enable an accurate modeling of the image acquisition geometry, we provided a step-by-step description of the 3D-to-3D VO software that is used to retrieve a MLE of the rover's location and attitude.

The software was first tested by processing two consecutive pairs of rectified images captured by the rover Opportunity of the NASA Mars Exploration Rovers mission. The maximum-likelihood motion estimate obtained by the algorithm allowed to detect and to mitigate errors in the SPICE mission kernels, which also caused inconsistencies in the archived point-clouds.

Our VO algorithm was then used to reconstruct Perseverance's location and orientation along several traverses on sols 65, 66 and 72 by processing archived raw stereo images. To quantify the differences between our VO-estimated and the telemetered paths, position discrepancies between the two solutions were computed. The accumulation of position differences along the drives can be partially explained by the presence of errors in the encoders-based position estimates that are compensated in our solution. Drive steps smaller than 1 m are fully consistent with telemetry data. The discrepancies inflate in correspondence to larger drive steps. Because of sparse tracked features in stereo pairs acquired after long distance travelled by the rover, the pose estimation is significantly affected by the measurement weighting of MLE filter.

To assess the accuracies of the two solutions, a method was implemented, based on the triangulation of common world-points observed before and after the entire path. By using our VO estimates to predict the landmarks coordinates at the end of the traverse, smaller discrepancies are obtained compared to the telemetry-based paths. The results support higher localization accuracies provided by our solution, which yields a reduction of the landmarks localization errors by a factor >2. A set of stereo pairs acquired on sol 120 from the same spot were also processed to assess the capabilities of our algorithm to deal with large rotations. The results indicate limited accumulation of position errors (<1.5 cm), and a reconstructed orientation in full agreement with the accurate IMU measurements (<0.1° errors).

To compare the localization performances of different VO-based localization methods, we processed the sequence of images acquired on sol 72 by also using a 3D-to-2D VO scheme (Supplementary Section 3). For drive steps smaller than 1 m, the 3D-to-3D and 3D-to-2D VO approaches produce comparable results. For longer motion steps, the 3D-to-2D VO method produces estimates that are more consistent with the refined telemetry data (<10 cm discrepancies) compared to the 3D-to-3D VO-based solution. This behavior may be partially explained by the uneven distribution of the tracked image-keypoints in case of longer drive steps, which are mostly related to the farther landmarks. The 3D-to-2D VO method tends to classify the few 3D-2D correspondences associated with the closer landmarks as outliers, since they yield the major reprojection errors. By excluding these data, the 3D-to-2D VO motion estimates rely on the farther landmarks only, limiting the adjustment to the onboard estimated motion (*i.e.*, retrieved from the image metadata), which is indeed used to track the keypoints between adjacent stereo pairs. On the other hand, the 3D-to-3D VO method preserves the information associated with the closer landmarks by weighting them more, leading to major discrepancies with respect to the telemetered motion.

VO represents a key localization technique that enables accurate updates of the pose of moving assets in the operational environment, although localization error will eventually accumulate over long traverses. To support safe navigation operations on demanding terrains (Gargiulo et al., 2021a), data-fusion approaches can be adopted to process additional measurements, such as sun sensors and star-trackers observations, and LiDAR data (Carle et al., 2010; Carle & Barfoot, 2010; Gargiulo et al., 2021b). Local bundle adjustment techniques can be also employed to reduce the drift of the reconstructed trajectory by jointly refining the position of the rover and the landmarks locations, enabling an accurate extended motion estimation. Furthermore, global localization accuracies in the operational environment can be attained by using correlation-based techniques based on Digital

Terrain Models (DTM). For example, local 3D maps of the vehicle's neighborhoods (retrieved from the stereo imaging data collected by the rover) can be compared to global high-resolution DTMs that are extracted from orbital stereo-images (Fergason et al., 2020) or by processing altimetry data (*e.g.*, MOLA (Smith et al., 2001); Genova, 2020), allowing for a further mitigation of the rover's pose drift errors after extended traverses. A combined processing of these datasets will enable significant enhancements in the trajectory reconstruction, especially after long drive steps without intermediate acquisitions of stereo images.

REFERENCES

Andolfo, S., Petricca, F., & Genova, A. (2021). Rovers Localization by using 3D-to-3D and 3D-to-2D Visual Odometry. *Proceedings of 2021 IEEE 8th International Workshop on Metrology for AeroSpace (MetroAeroSpace)*, 334–339. https://doi.org/10.1109/MetroAeroSpace51421.2021.9511741

Andolfo, S., Petricca, F., & Genova, A. (2022). Visual Odometry analysis of the NASA Mars 2020 Perseverance rover's images. *Proceedings of 2022 IEEE 9th International Workshop on Metrology for AeroSpace (MetroAeroSpace)*.

Arun, K. S., Huang, T. S., & Blostein, S. D. (1987). Least-squares fitting of two 3-D point sets. IEEE Transactions on Pattern Analysis and Machine Intelligence, 9(5), 698–700. https://doi.org/10.1109/TPAMI.1987.4767965

Bay, H., Tuytelaars, T., & Van Gool, L. (2006). SURF: Speeded Up Robust Features. In A. Leonardis, H. Bischof, A. Pinz (Eds.), *Lecture Notes in Computer Science: Vol. 3951. Computer Vision – ECCV 2006* (pp. 404–417). Springer. https://doi.org/10.1007/11744023_32

Bailey, T., Nieto, J., Guivant, J., Stevens, M., & Nebot, E. (2006). Consistency of the EKF-SLAM Algorithm. *Proceedings of the 2006 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 3562–3568. https://doi.org/10.1109/IROS.2006.281644

Biesiadecki, J. J., & Maimone, M. (2006). The Mars Exploration Rover surface mobility flight software driving ambition. *Proceedings of 2006 IEEE Aerospace Conference*. https://doi.org/10.1109/AERO.2006.1655723

Brown, D. C. (1971). Close-range camera calibration. Photogrammetric Engineering, 37, 855–866. https://www.asprs.org/wp-content/uploads/pers/1971journal/aug/1971_aug_855-866.pdf

Carle, P. J. F., Furgale, P. T., & Barfoot, T. D. (2010). Long-range rover localization by matching LIDAR scans to orbital elevation maps. Journal of Field Robotics, 27, 344–370. https://doi.org/10.1002/rob.20336

Carle, P. J. F., & Barfoot, T. D. (2010). Global rover localization by matching lidar and orbital 3D maps. *Proceedings of 2010 IEEE International Conference on Robotics and Automation*, 881–886. https://doi.org/10.1109/ROBOT.2010.5509767

Chen, A. (2014). Mars Exploration Rover Project Software Interface Specification (SIS): Camera Experiment Data Record (EDR) and Reduced Data Record (RDR) Operations and Science Data Products, Version 4.4, JPL D-22846, July 31, 2014. https://pds-imaging.jpl.nasa.gov/data/mer/opportunity/mer1no_0xxx/document/CAMSIS_latest.PDF

Corke, P., Strelow, D., & Singh, S. (2004). Omnidirectional visual odometry for a planetary rover. *Proceedings of the 2004 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) (IEEE Cat. No.04CH37566)*, 4007–4012. https://doi.org/10.1109/IROS.2004.1390041

Deen, R. (2022). Mars 2020 Project Software Interface Specification (SIS): Rover PLACES data products for PDS, Version 1.0, February 02, 2022. https://pds-geosciences.wustl.edu/m2020/urn-nasa-pds-mars2020_rover_places/document/Mars2020_Rover_PLACES_PDS_SIS.pdf

Di, K., & Li, R. (2004). CAHVOR camera model and its photogrammetric conversion for planetary applications. Journal of Geophysical Research, 109(E4). https://doi.org/10.1029/2003JE002199

Di, K., Xu, F., Wang, J., Agarwal, S., Brodyagina, E., Li, R., & Matthies, L. (2008). Photogrammetric processing of rover imagery of the 2003 Mars Exploration Rover mission. ISPRS Journal of Photogrammetry and Remote Sensing, 63(2), 181–201. https://doi.org/10.1016/j.isprsjprs.2007.07.007

Enright, J., Barfoot, T., & Soto, M. (2012). Star tracking for planetary rovers. *Proceedings of 2012 IEEE Aerospace Conference*, 1–13. https://doi.org/10.1109/AERO.2012.6187042

Farley, K. A., Williford, K. H., Stack, K. M., Bhartia, R., Chen, A., de la Torre, M., Hand, K., Goreva, Y., Herd, C. D. K., Hueso, R., Liu, Y., Maki, J. N., Martinez, G., Moeller, R. C., Nelessen, A., Newman, C. E., Nunes, D., Ponce, A., Spanovich, N., ..., Wiens, R. C. (2020). Mars 2020 Mission Overview. Space Science Reviews, 216. https://doi.org/10.1007/s11214-020-00762-y

Fasogbon, P., & Aksu, E. (2019). Calibration of Fisheye Camera Using Entrance Pupil. *Proceedings of 2019 IEEE International Conference on Image Processing (ICIP)*, 469–473. https://doi.org/10.1109/ICIP.2019.8803832

Fergason, R. L., Hare, T. M., Mayer, D. P., Galuszka, D. M., Redding, B. L., Smith, E. D., Shinaman, J. R., Cheng, Y., & Otero, R. E. (2020, March). Mars 2020 Terrain Relative Navigation Flight Product Generation: Digital Terrain Model and Orthorectified Image Mosaics. In *51st Lunar and Planetary Science Conference (LPSC)*. https://www.hou.usra.edu/meetings/lpsc2020/pdf/2020.pdf

Ferrera, M., Moras, J., Trouvé-Peloux P., & Creuze, V. (2019). Real-Time Monocular Visual Odometry for Turbid and Dynamic Underwater Environments. Sensors, 19(3):687. https://doi.org/10.3390/s19030687

Gargiulo, A. M., Di Stefano, I., & Genova, A. (2021). Model-Based Slippage Estimation to Enhance Planetary Rover Localization with Wheel Odometry. Applied Sciences, 11(12):5490. https://doi.org/10.3390/app11125490

Gargiulo, A. M., Di Stefano, I., & Genova, A. (2021). Numerical Simulations for Planetary Rovers Safe Navigation and LIDAR Based Localization. *Proceedings of 2021 IEEE 8th International Workshop on Metrology for AeroSpace (MetroAeroSpace)*, 418–423. https://doi.org/10.1109/MetroAeroSpace51421.2021.9511774

Gennery, D. B. (2001). Least-Squares Camera Calibration Including Lens Distortion and Automatic Editing of Calibration Points. In A. Gruen A., T. S. Huang (Eds.), *Springer Series in Information Sciences: Vol. 34. Calibration and Orientation of Cameras in Computer Vision* (pp. 123–136). Springer. https://doi.org/10.1007/978-3-662-04567-1_5

Gennery, D. B. (2006). Generalized Camera Calibration Including Fish-Eye Lenses. International Journal of Computer Vision, 68(3), 239–266. https://doi.org/10.1007/s11263-006-5168-1

Genova, A. (2020). ORACLE: A mission concept to study Mars' climate, surface and interior. Acta Astronautica, 166, 317–329. https://doi.org/10.1016/j.actaastro.2019.10.006

Giubilato, R., Stürzl, W., Wedler, A., & Triebel, R. (2022). Challenges of SLAM in Extremely Unstructured Environments: The DLR Planetary Stereo, Solid-State LiDAR, Inertial Dataset. IEEE Robotics and Automation Letters, 7(4), 8721–8728. https://doi.org/10.1109/LRA.2022.3188118

Gonzalez, R., & Iagnemma, K. (2018). Slippage estimation and compensation for planetary exploration rovers. State of the art and future challenges. Journal of Field Robotics, 35, 564– 577. https://doi.org/10.1002/rob.21761

Harris, C. G., & Stephens, M. J. (1988). A Combined Corner and Edge Detector. In C. J. Taylor (Eds.), *Proceedings of the Alvey Vision Conference, AVC 1988, Manchester, UK, September, 1988* (pp. 23.1–23.6). Alvey Vision Club. https://doi.org/10.5244/C.2.23

Hayes, A. G., Corlies, P., Tate, C., Barrington, M., Bell, J. F., Maki, J. N., Caplinger, M., Ravine, M., Kinch, K. M., Herkenhoff, K., Horgan, B., Johnson, J., Lemmon, M., Paar, G., Rice, M. S., Jensen, E., Kubacki, T. M., Cloutis, E., Deen, R., …, Winchell, K. (2021). Pre-Flight Calibration of the Mars 2020 Rover Mastcam Zoom (Mastcam-Z) Multispectral, Stereoscopic Imager. Space Science Reviews, 217, 29. https://doi.org/10.1007/s11214-021-00795-x

Hidalgo-Carrió, J., Poulakis, P., & Kirchner, F. (2018). Adaptive localization and mapping with application to planetary rovers. Journal of Field Robotics, 35, 961–987. https://doi.org/10.1002/rob.21790

Howard, A. (2008). Real-time stereo visual odometry for autonomous ground vehicles. *Proceedings of 2008 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 3946–3952. https://doi.org/10.1109/IROS.2008.4651147

Johnson, A. E., Goldberg, S. B., Cheng, Y., & Matthies, L. H. (2008). Robust and Efficient Stereo Feature Tracking for Visual Odometry. *Proceedings of the 2008 IEEE International Conference on Robotics and Automation*, 39–46. https://doi.org/10.1109/ROBOT.2008.4543184

Kannala, J., & Brandt, S. S. (2006). A generic camera model and calibration method for conventional, wide-angle, and fish-eye lenses. IEEE Transactions on Pattern Analysis and Machine Intelligence, 28, 1335–1340. https://doi.org/10.1109/TPAMI.2006.153

Kim, P., Lee, H., & Kim, H. J. (2019). Autonomous flight with robust visual odometry under dynamic lighting conditions. Autonomous Robots, 43, 1605–1622. https://doi.org/10.1007/s10514-018-9816-4

Ma, Y., Liu, S., Sima, B., Wen, B., Peng, S., & Jia, Y. (2020). A precise visual localisation method for the Chinese Chang'e-4 Yutu-2 rover. The Photogrammetric Record, 35(169), 10–39. https://doi.org/10.1111/PHOR.12309

Maimone, M., Cheng, Y., & Matthies, L. (2007). Two Years of Visual Odometry on the Mars Exploration Rovers. Journal of Field Robotics, 24(3), 169–186. https://doi.org/10.1002/rob.20184

Maki, J. N., Bell III, J. F., Herkenhoff, K. E., Squyres, S. W., Kiely, A., Klimesh, M., Schwochert, M., Litwin, T., Willson, R., Johnson, A., Maimone, M., Baumgartner, E., Collins, A., Wadsworth, M., Elliot, S. T., Dingizian, A., Brown, D., Hagerott, E. C., Scherr, L., …, Lorre, J. (2003). Mars Exploration Rover Engineering Cameras, Journal of Geophysical Research, 108(E12), 8071. https://doi.org/10.1029/2003JE002077

Maki, J. N., Gruel, D., McKinney, C., Ravine, M. A., Morales, M., Lee, D., Willson, R., Copley-Woods, D., Valvo, M., Goodsall, T., McGuire, J., Sellar, R. G., Schaffner, J. A., Caplinger, M. A., Shamah, J. M., Johnson, A. E., Ansari, H., Singh, K., Litwin, T., ..., Algermissen, S. (2020). The Mars 2020 Engineering Cameras and Microphone on the Perseverance Rover: A Next-Generation Imaging System for Mars Exploration. Space Science Reviews, 216(8), 1–48. https://doi.org/10.1007/s11214-020-00765-9

Matthies, L., & Shafer, S. (1987). Error modeling in stereo navigation. IEEE Journal on Robotics and Automation 3(3), 239–248, June 1987. https://doi.org/10.1109/JRA.1987.1087097

Matthies, L., Maimone, M., Johnson, A., Cheng, Y., Willson, R., Villalpando, C., Goldberg, S., Huertas, A., Stein A., & Angelova, A. (2007). Computer Vision on Mars. International Journal of Computer Vision, 75, 67–92. https://doi.org/10.1007/s11263-007-0046-z

Mishkin, A. H., Morrison, J. C., Nguyen, T. T., Stone, H. W., Cooper, B. K., & Wilcox, B. H. (1998). Experiences with operations and autonomy of the Mars Pathfinder Microrover. *Proceedings of the 1998 IEEE Aerospace Conference* (Cat. No.98TH8339), 337–351. https://doi.org/10.1109/AERO.1998.687920

Nistér, D., Naroditsky, O., & Bergen, J. (2004). Visual odometry. *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2004)*. https://doi.org/10.1109/CVPR.2004.1315094

Olson, C. F., Matthies, L. H., Schoppers, M., & Maimone, M. W. (2003). Rover navigation using stereo ego-motion. Robotics and Autonomous Systems, 43(4), 215–229. https://doi.org/10.1016/S0921-8890(03)00004-6

Pertile, M., Chiodini, S., & Debei, S. (2014). Comparison of visual odometry systems suitable for planetary exploration. *Proceedings of 2014 Metrology for AeroSpace (MetroAeroSpace)*, 232–237. https://doi.org/10.1109/MetroAeroSpace.2014.6865926

Rankin, A., Maimone, M., Biesiadecki, J., Patel, N., Levine, D., & Toupet, O. (2020). Driving Curiosity: Mars Rover Mobility Trends During the First Seven Years. *Proceedings of 2020 IEEE Aerospace Conference*, 1–19. https://doi.org/10.1109/AERO47225.2020.9172469

Ruoff, N., Deen, R., Pariser, O. (2021). Mars2020 Project Software Interface Specification (SIS): Camera Data products, Version 3.1, JPL D-99960, February 22, 2022. https://pds-geosciences.wustl.edu/m2020/urn-nasa-pds-mars2020_mission/document_camera/Mars2020_Camera_SIS.pdf

Scaramuzza, D., Fraundorfer, F., & Siegwart, R. (2009). Real-time monocular visual odometry for on-road vehicles with 1-point RANSAC. *Proceedings of 2009 IEEE International Conference on Robotics and Automation (ICRA 2009)*, 4293–4299. https://doi.org/10.1109/ROBOT.2009.5152255

Smith, D. E., Zuber, M. T., Frey, H. V., Garvin, J. B., Head, J. W., Muhleman, D. O., Pettengill, G. H., Phillips, R. J., Solomon, S. C., Zwally, H. J., Banerdt, W. B., Duxbury, T. C., Golombek, M. P., Lemoine, F. G., Neumann, G. A., Rowlands, D. D., Aharonson, O., Ford, P. G., & Ivanov A. B. (2001). Mars Orbiter Laser Altimeter: Experiment summary after the first year of global mapping of Mars. Journal of Geophysical Research, 106(E10), 23689–23722. https://doi.org/10.1029/2000JE001364

So, E. W. Y., Yoshimitsu, T., & Kubota, T. (2011). Visual Odometry for a Hopping Rover on an Asteroid Surface using Multiple Monocular Cameras. Advanced Robotics, 25(6–7), 893–921. https://doi.org/10.1163/016918611X563355

Teixeira, B., Silva, H., Matos, A., & Silva, E. (2020). Deep Learning for Underwater Visual Odometry Estimation. IEEE Access, 8, 44687–44701. https://doi.org/10.1109/ACCESS.2020.2978406

Townson, D., Woods, M., & Carnochan, S. (2018, June). ExoMars VisLoc – The industrialised, visual localisation system for the ExoMars rover. In *International Symposium on Artificial Intelligence, Robotics and Automation in Space (i-SAIRAS). https://robotics.estec.esa.int/i-SAIRAS/isairas2018/Papers/Session%207a/4_129Townson_Woods.pdf*

Utz, H., & Fluckiger, L. (2021). Rover Flight Software for the VIPER Mission. In *15th Annual Flight Software Workshop*. https://ntrs.nasa.gov/api/citations/20210025416/downloads/VIPER-RSW-FSW22-HUtz-et-al-v5.pdf

Verma, V. (October 29, 2021). Driving Farther and Faster With Autonomous Navigation and Helicopter Scouting. *Blog: Mars Perseverance Rover Mission.* https://mars.nasa.gov/mars2020/mission/status/342/driving-farther-and-faster-with-autonomous-navigation-and-helicopter-scouting/

Verma, V., Hartman, F., Rankin, A., Maimone, M., Del Sesto, T., Toupet, O., Graser, E., Myint, S., Davis, K., Klein, D., Koch, J., Brooks, S., Bailey, P., Justice, H., Dolci, M., & Ono, H. (2022). First 210 solar days of Mars Perseverance Robotic Operations - Mobility, Robotics Arm, Sampling, and Helicopter. *Proceedings of 2022 IEEE Aerospace Conference*. https://www-robotics.jpl.nasa.gov/media/documents/ro_final_v1.pdf

Winter, M., Barclay, C., Pereira, V., Lancaster, R., Caceres, M., McManamon, K., Nye, B., Silva, N., Lachat, D., & Campana, M. (2015, May). ExoMars Rover Vehicle: detailed description of the GNC system. In *13th Symposium on Advanced Space Technologies in Robotics and Automation (ASTRA)*. https://robotics.estec.esa.int/ASTRA/Astra2015/Papers/Session%201A/96038_Winter.pdf

Witte, I. R., Bekker, D. L., Chen, M. H., Criss, T. B., Jenkins, S. N., Mehta, N. L., Sawyer, C. A., Stipes, J. A., & Thomas, J. R. (2019). No GPS? No Problem! Exploring the Dunes of Titan with Dragonfly Using Visual Odometry. *Proceedings of the AIAA Scitech 2019 Forum*. https://doi.org/10.2514/6.2019-1177

Wudenka, M., Müller, M. G., Demmel, N., Wedler, A., Triebel, R., Cremers, D., & Stürzl, W. (2021). Towards Robust Monocular Visual Odometry for Flying Robots on Planetary Missions. *Proceedings of 2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 8737–8744. https://doi.org/10.1109/IROS51168.2021.9636844

Yakimovsky, Y., & Cunningham, R. (1978). A system for extracting three-dimensional measurements from a stereo pair of TV cameras. Computer graphics and image processing, 7(2), 195– 210. https://doi.org/10.1016/0146-664X(78)90112-0

Young, M. (1971). Pinhole Optics. Applied Optics, 10(12), 2763–2767. https://doi.org/10.1364/AO.10.002763

Zhu, Q., Wu, B., & Wan, N. (2007). A sub-pixel location method for interest points by means of the Harris interest strength. The Photogrammetric Record, 22(120), 321–335. https://doi.org/10.1111/j.1477-9730.2007.00450.x

APPENDIX A

In this section, we outline how to compute the Jacobian matrix associated with the partial derivatives of the triangulated coordinates of a world-point, $\boldsymbol{P}$, with respect to the associated left and right image-points 2D coordinates. We present the computations needed to retrieve the first column of the Jacobian matrix, which is related to the partial derivative of $\boldsymbol{P}$ with respect to the x-coordinate of the left pixel. The other columns of the Jacobian matrix can be retrieved performing analogous computational steps.

Since the 3-dimensional vector $\boldsymbol{P}$ depends on x_L through $\boldsymbol{P}_L$ and $\boldsymbol{P}_R$ (see Eq. (11)), it results that:

$$\frac{\partial \boldsymbol{P}}{\partial x_L} = \frac{1}{2}\left[\frac{\partial \boldsymbol{P}_L}{\partial x_L} + \frac{\partial \boldsymbol{P}_R}{\partial x_L}\right] \qquad (\mathrm{A.1})$$

By applying the chain rule for the partial derivatives to the right-hand side terms of Eq. (A.1), one obtains:

$$\frac{\partial \boldsymbol{P}_i}{\partial x_L} = \frac{\partial \boldsymbol{P}_i}{\partial \boldsymbol{C}_i'}\frac{\partial \boldsymbol{C}_i'}{\partial x_L} + \frac{\partial \boldsymbol{P}_i}{\partial \boldsymbol{r}_i}\frac{\partial \boldsymbol{r}_i}{\partial x_L} + \frac{\partial \boldsymbol{P}_i}{\partial m_i}\frac{\partial m_i}{\partial x_L} \quad i = L, R \tag{A.2}$$

The partial derivatives of $\boldsymbol{P}_i$ with respect to $\boldsymbol{C}_i'$, $\boldsymbol{r}_i$ and m_i $(i = L, R)$ and the partial derivatives of m_i with respect to $\boldsymbol{C}_i'$ and $\boldsymbol{r}_i$ $(i = L, R)$ are computed by applying the chain rule to differentiate Eqs. (7) and Eqs. (10), respectively. The partial derivatives of $\boldsymbol{C}_i'$ and $\boldsymbol{r}_i$ $(i = L, R)$ with respect to the 2D coordinates of the image-keypoints are finally retrieved by differentiating the full set of nonlinear equations based on the CAHVORE camera parameters. A thorough description of the main steps required to retrieve these quantities is provided hereafter (to simplify the notation, the subscript i that refers to left/right quantities is not reported explicitly).

First, the adjusted (*i.e.*, distortion compensated) viewing ray direction $\boldsymbol{r}$ is computed. Given the (x, y)-coordinates of a keypoint, the associated viewing ray is first projected out (from the image 2D space into 3D world space) according to the CAHV model (that neglects distortion and the entrance pupil displacement) from the point $\boldsymbol{C}'$ (*i.e.*, the adjusted location of the entrance pupil) as follows:

$$\boldsymbol{r}' = \frac{(\boldsymbol{V} - y\boldsymbol{A}) \times (\boldsymbol{H} - x\boldsymbol{A})}{\boldsymbol{A} \cdot \boldsymbol{V} \times \boldsymbol{H}} \tag{A.3}$$

where $\boldsymbol{A}, \boldsymbol{H}$ and $\boldsymbol{V}$ are the *axis*, *horizontal* and *vertical* vectors associated with the CAHVORE model. The vector $\boldsymbol{r}'$ can be written as the sum of two vectors, one parallel to the optical axis $\boldsymbol{O}$, and the other along the direction $\widehat{\boldsymbol{\lambda}'}$, which is orthogonal to $\boldsymbol{O}$, as follows:

$$\boldsymbol{r}' = \zeta'\boldsymbol{O} + \lambda'\widehat{\boldsymbol{\lambda}'} = \zeta'(\boldsymbol{O} + \chi'\widehat{\boldsymbol{\lambda}'}) \tag{A.4}$$

where $\chi' = \lambda'/\zeta'$. The effect of the radial distortion is modeled as an apparent shift of the 3D point associated with the image keypoint (x, y) in a direction orthogonal to $\boldsymbol{O}$ by an amount $\mu\lambda$, being $\lambda = \boldsymbol{r} \cdot \widehat{\boldsymbol{\lambda}'}$, and μ the distortion polynomial defined as:

$$\mu = \boldsymbol{R}(1) + \boldsymbol{R}(2)\chi^2 + \boldsymbol{R}(3)\chi^4 \tag{A.5}$$

where $\boldsymbol{R}(j)$ denotes the j^{th} component $(j = 1,2,3)$ of the radial vector $\boldsymbol{R}$, and the parameter $\chi = f(\theta)$ is a function of the off-axis angle θ between $\boldsymbol{r}$ and $\boldsymbol{O}$. Therefore, it results that $\lambda' = (1 + \mu)\lambda$; similarly, the parameter χ' can be expressed as:

$$\chi' = (1 + \mu)\chi = \big(1 + \boldsymbol{R}(1)\big)\chi + \boldsymbol{R}(2)\chi^3 + \boldsymbol{R}(3)\chi^5 \tag{A.6}$$

Once the apparent 3D displacement of the observed point is compensated, the adjusted viewing unit vector can be written as:

$$\boldsymbol{r} = \sin(\theta)\widehat{\boldsymbol{\lambda}'} + \cos(\theta)\boldsymbol{O} \tag{A.7}$$

To compute the off-axis angle θ, the nonlinear equation (A.6) is solved for χ by using Newton-Raphson method (with χ' as the initial approximation for χ), and then $\chi = f(\theta)$ is solved for θ (for ideal fisheye lenses, $\chi = \theta$). The ray $\boldsymbol{r}$ is considered projected from the adjusted entrance pupil location $\boldsymbol{C}'$, which is computed according to (1).

The partial derivatives of $\boldsymbol{C}'$ and $\boldsymbol{r}$ with respect to the x-coordinate of the image keypoint are retrieved by applying the chain rule to Eq. (1) and Eq. $(\mathrm{A.}\,7)$, as follows:

$$\begin{gathered}
\frac{\partial \boldsymbol{C}'}{\partial x} = \frac{\partial \boldsymbol{C}'}{\partial \boldsymbol{r}'}\frac{\partial \boldsymbol{r}'}{\partial x} = \left(\frac{\partial \boldsymbol{C}'}{\partial \boldsymbol{C}}\frac{\partial \boldsymbol{C}}{\partial \boldsymbol{r}'} + \frac{\partial \boldsymbol{C}'}{\partial \boldsymbol{O}}\frac{\partial \boldsymbol{O}}{\partial \boldsymbol{r}'} + \frac{\partial \boldsymbol{C}'}{\partial s}\frac{\partial s}{\partial \boldsymbol{r}'}\right)\frac{\partial \boldsymbol{r}'}{\partial x} \\
\frac{\partial \boldsymbol{r}}{\partial x} = \frac{\partial \boldsymbol{r}}{\partial \boldsymbol{r}'}\frac{\partial \boldsymbol{r}'}{\partial x} = \left(\frac{\partial \boldsymbol{r}}{\partial \widehat{\boldsymbol{\lambda}}'}\frac{\partial \widehat{\boldsymbol{\lambda}}'}{\partial \boldsymbol{r}'} + \frac{\partial \boldsymbol{r}}{\partial \theta}\frac{\partial \theta}{\partial \boldsymbol{r}'} + \frac{\partial \boldsymbol{r}}{\partial \boldsymbol{O}}\frac{\partial \boldsymbol{O}}{\partial \boldsymbol{r}'}\right)\frac{\partial \boldsymbol{r}'}{\partial x}
\end{gathered} \tag{A.8}$$

where $\partial \boldsymbol{r}'/\partial x$ is retrieved from eq. $(\mathrm{A.}\,3)$, and the partial derivatives of $\boldsymbol{C}'$ and $\boldsymbol{r}$ with respect to $\boldsymbol{r}'$ are computed accordingly to (Gennery, 2006):

$$\begin{aligned}
&\frac{\partial \boldsymbol{C}}{\partial \boldsymbol{r}'} = \frac{\partial \boldsymbol{C}'}{\partial s}\frac{\partial s}{\partial \boldsymbol{r}'} = \boldsymbol{O}\frac{\partial s}{\partial \theta}\frac{\partial \theta}{\partial \boldsymbol{r}'} \\
&\frac{\partial \boldsymbol{r}}{\partial \boldsymbol{r}'} = \frac{\partial \boldsymbol{r}}{\partial \widehat{\boldsymbol{\lambda}}'}\frac{\partial \widehat{\boldsymbol{\lambda}}'}{\partial \boldsymbol{r}'} + \frac{\partial \boldsymbol{r}}{\partial \theta}\frac{\partial \theta}{\partial \boldsymbol{r}'} = \frac{\sin(\theta)}{\lambda'}\left(\mathbf{I} - \widehat{\boldsymbol{\lambda}}'\widehat{\boldsymbol{\lambda}}'^{\mathrm{T}} - \boldsymbol{O}\boldsymbol{O}^{\mathrm{T}}\right) + \left(\cos(\theta)\widehat{\boldsymbol{\lambda}}' - \sin(\theta)\boldsymbol{O}\right)\frac{\partial \theta}{\partial \boldsymbol{r}'} \\
&\frac{\partial \theta}{\partial \boldsymbol{r}'} = \frac{\partial \theta}{\partial \chi}\frac{\partial \chi}{\partial \chi'}\frac{\partial \chi'}{\partial \boldsymbol{r}'}
\end{aligned} \tag{A.9}$$

Analogous computations based on Eqs. $(\mathrm{A.}\,8 - \mathrm{A.}\,9)$ can be carried out to compute the remaining partial derivatives that are required to define the Jacobian (*i.e.*, the partial derivatives of the left/right $\boldsymbol{C}'$ and $\boldsymbol{r}$ with respect to the x-coordinate and the y-coordinate of the left/right image keypoint associated with the landmark $\boldsymbol{P}$). To speed up the computations, some terms can be neglected, since the left entrance pupil and viewing ray do not depend on the coordinates of the right corner-point, and vice-versa, leading to:

$$\frac{\partial \boldsymbol{C}'_L}{\partial x_R} = \frac{\partial \boldsymbol{C}'_L}{\partial y_R} = \frac{\partial \boldsymbol{r}_L}{\partial x_R} = \frac{\partial \boldsymbol{r}_L}{\partial y_R} = \frac{\partial \boldsymbol{C}'_R}{\partial x_L} = \frac{\partial \boldsymbol{C}'_R}{\partial y_L} = \frac{\partial \boldsymbol{r}_R}{\partial x_L} = \frac{\partial \boldsymbol{r}_R}{\partial y_L} = 0 \tag{A.10}$$

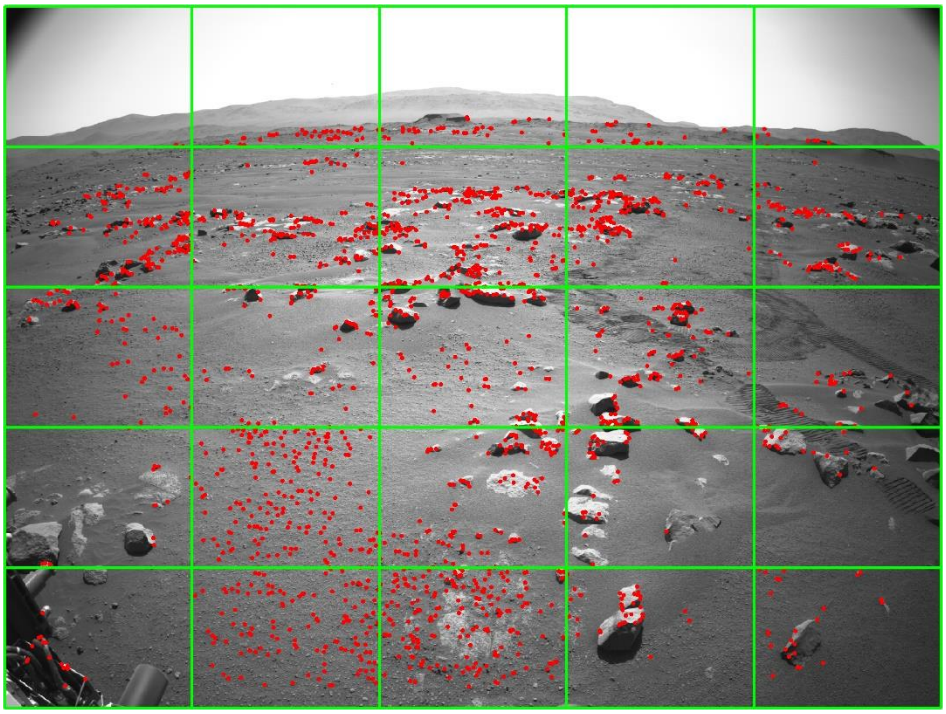

Figure 1. Detected corner points (**red**) in the left image of the first stereo pair ($\mathcal{L}_1$) acquired on sol 65. The Regions of Interest (ROI) are highlighted by green squares.

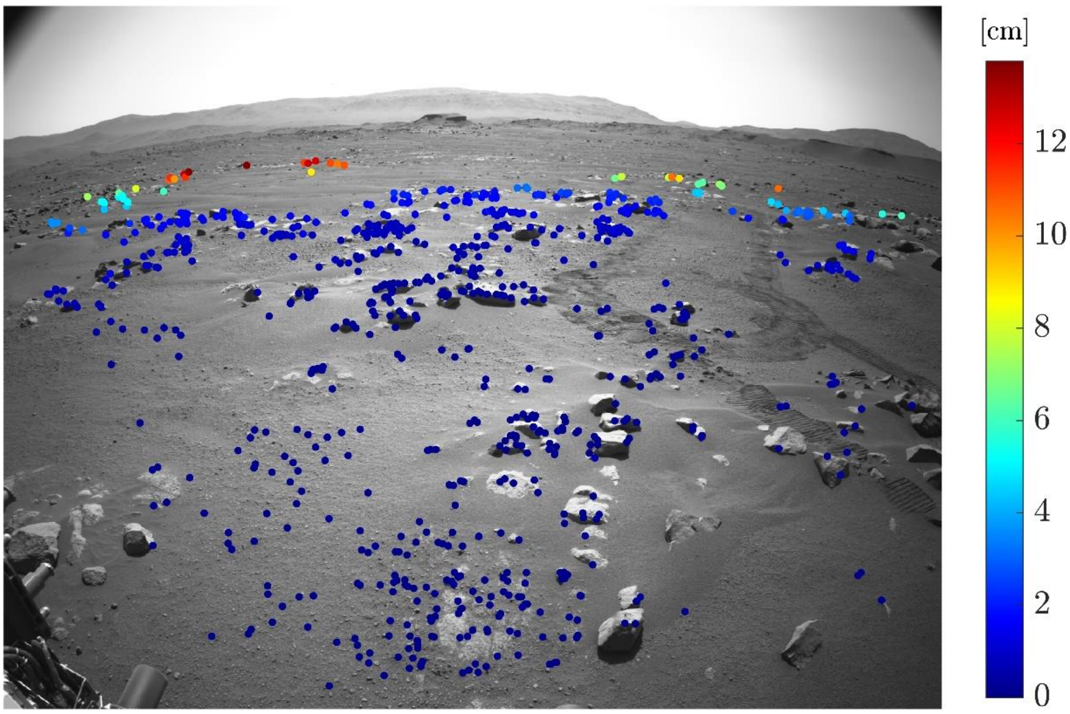

Figure 2. Triangulation error, represented through the length of the minimum distance segment joining the left and right viewing rays associated with a pair of matched corners. As expected, the farther landmarks are affected by larger triangulation errors. A maximum error of 15 cm is imposed to filter out unreliable landmarks. The left image of the first stereo pair ($\mathcal{L}_1$) acquired on sol 65 is shown on the background.

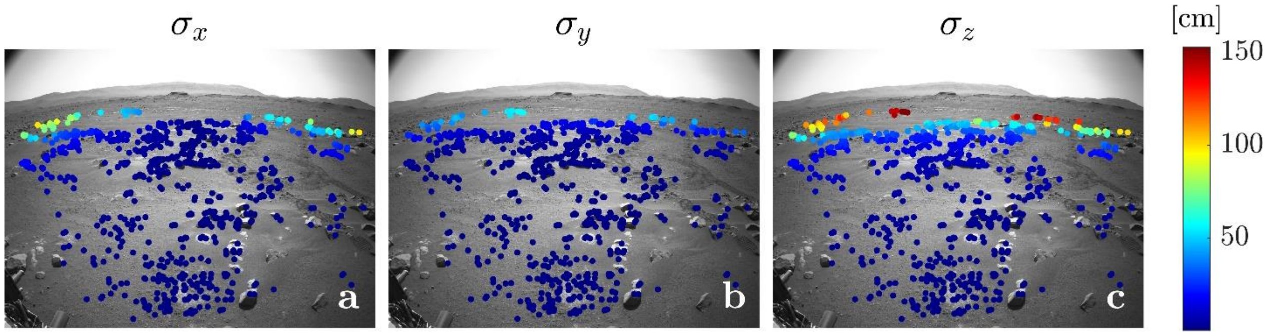

Figure 3. One standard deviation formal uncertainty of the triangulated landmarks coordinates, shown on the corner-points $\boldsymbol{p}_{\mathcal{L}_1}$ of the left image of the first stereo pair ($\mathcal{L}_1$) acquired on sol 65. The uncertainties are referred to the camera-centered frame (*i.e.*, Z-axis along the camera boresight; horizontal X-axis aligned with the image rows, from right to left; vertical Y-axis aligned with the image columns, from bottom to top, completing the right-hand triad). The uncertainties related to the X- (**a**), Y- (**b**), and Z-direction (**c**) all show a strong correlation with the relative distance of the landmarks from the rover (*i.e.*, farther landmarks are associated with greater triangulation uncertainties). The uncertainty distribution also depends on the orientation of the camera frame axes with respect to the line-of-sight direction (*i.e.*, the direction from the camera to the landmark, along which the 3D uncertainty distribution is elongated).

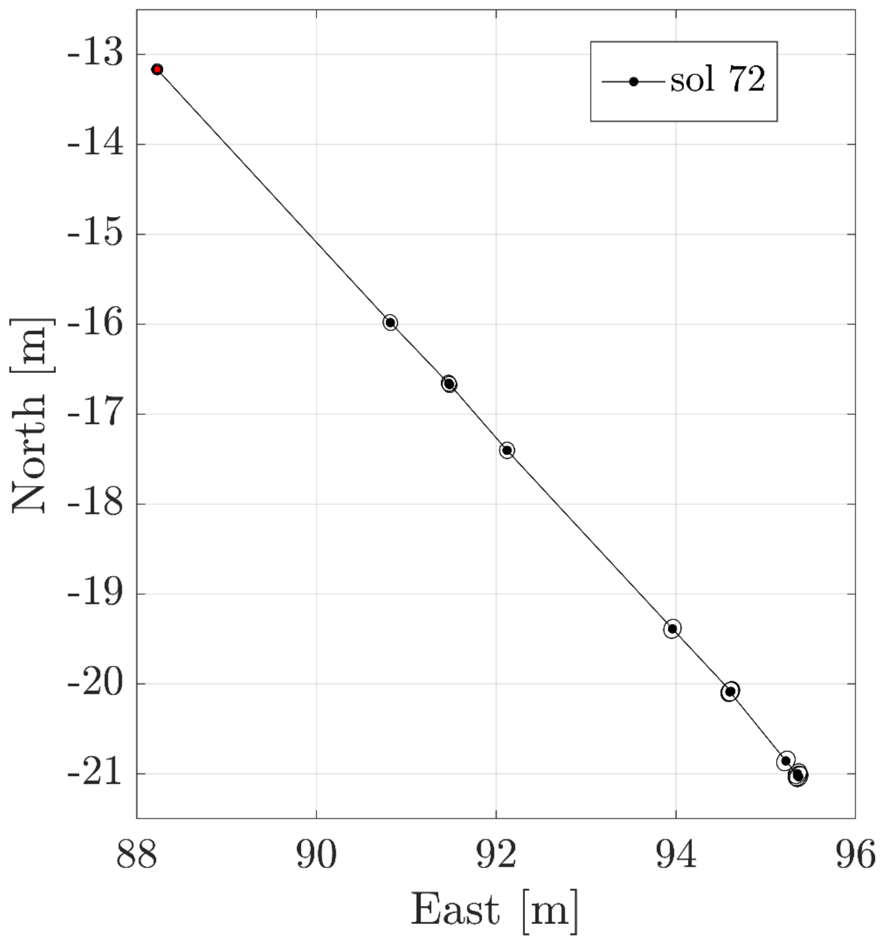

Figure 4. Reconstructed Perseverance's path on sol 72, and propagated 3-σ formal uncertainty ellipses. The rover's trajectory is recovered by using our VO algorithm, and the points where the ellipses are centered represent Perseverance's estimated locations at new stereo pair acquisitions. The red point represents Perseverance's initial position. The 3-σ ellipses are retrieved from the propagated rover's pose covariance ($\mathbf{\Sigma}_k$) that is computed by combining the rover's pose covariance at the previous step ($\mathbf{\Sigma}_{k-1}$) with the covariance associated with the estimated drive step ($\mathbf{\Sigma}_{k-1}^{k}$). The uncertainties related to the initial position and attitude vectors are assumed uncorrelated; the initial pose covariance is defined as $\mathbf{\Sigma}_0 = \mathrm{diag}(\sigma_x, \sigma_y, \sigma_z, \sigma_{\theta_x}, \sigma_{\theta_y}, \sigma_{\theta_z})$, with $\sigma_i = 2$ cm and $\sigma_{\theta_i} = 0.1°$ $(i = x, y, z)$.

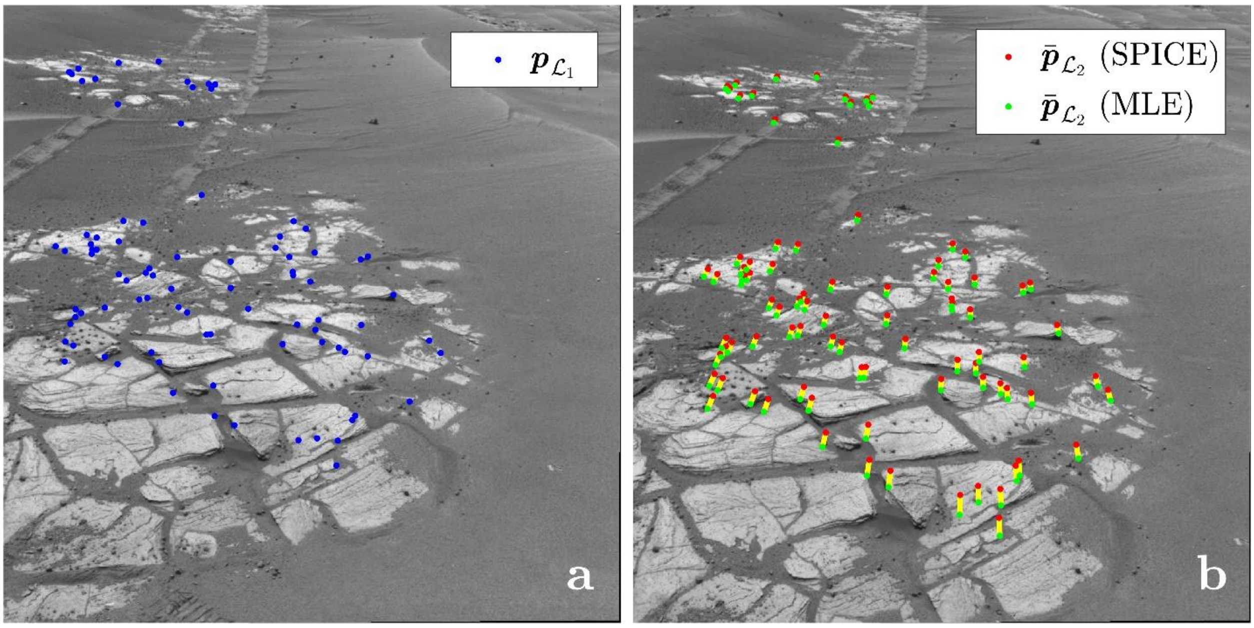

Figure 5. Left images acquired by Opportunity's left NavCam on sols 839-840 (5-6 June 2006), before (**a**) and after (**b**) a motion step. The left panel shows the left corner-points used for the first stereo-triangulation (**blue**), which provides the first 3D point-cloud (*i.e.*, triangulated landmarks before the drive step). The reprojection of the first point-cloud onto the second left image accordingly to the archived rover's position and attitude (*i.e.*, SPICE mission kernels) yields 2D image-points (**red**) that are not consistent with the corner-points detected in the first left image (**blue**), indicating errors on the rover's motion estimated onboard. A refined pose update is enabled by our maximum-likelihood VO motion estimate (MLE), which yields reprojected points (**green**) that are fully consistent with the landmarks observed before the drive step (**blue**). The reprojection vectors shown as yellow lines highlight the discrepancies between the two sets of reprojected points, which are retrieved by using the archived rover's motion and our pose estimate.

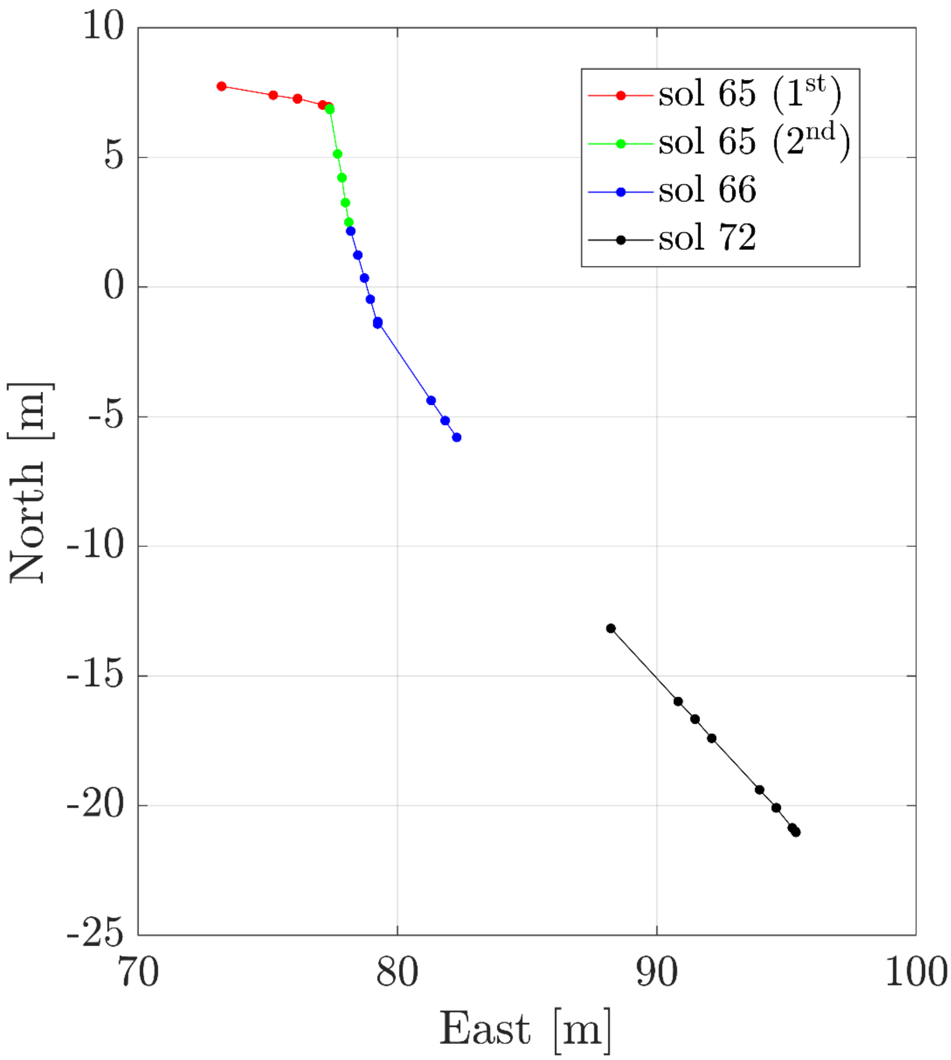

Figure 6. Perseverance's path on sol 65 (1st leg: **red**; 2nd leg: **green**), 66 (**blue**) and 72 (**black**) based on our VO solution. The initial rover's location for each leg is retrieved from telemetry data, and the points displayed along the path represent our estimated locations at new stereo pairs acquisitions. Perseverance's positions are referred to site 3, and are expressed in the *local level* (*i.e.*, North-East-Nadir) frame.

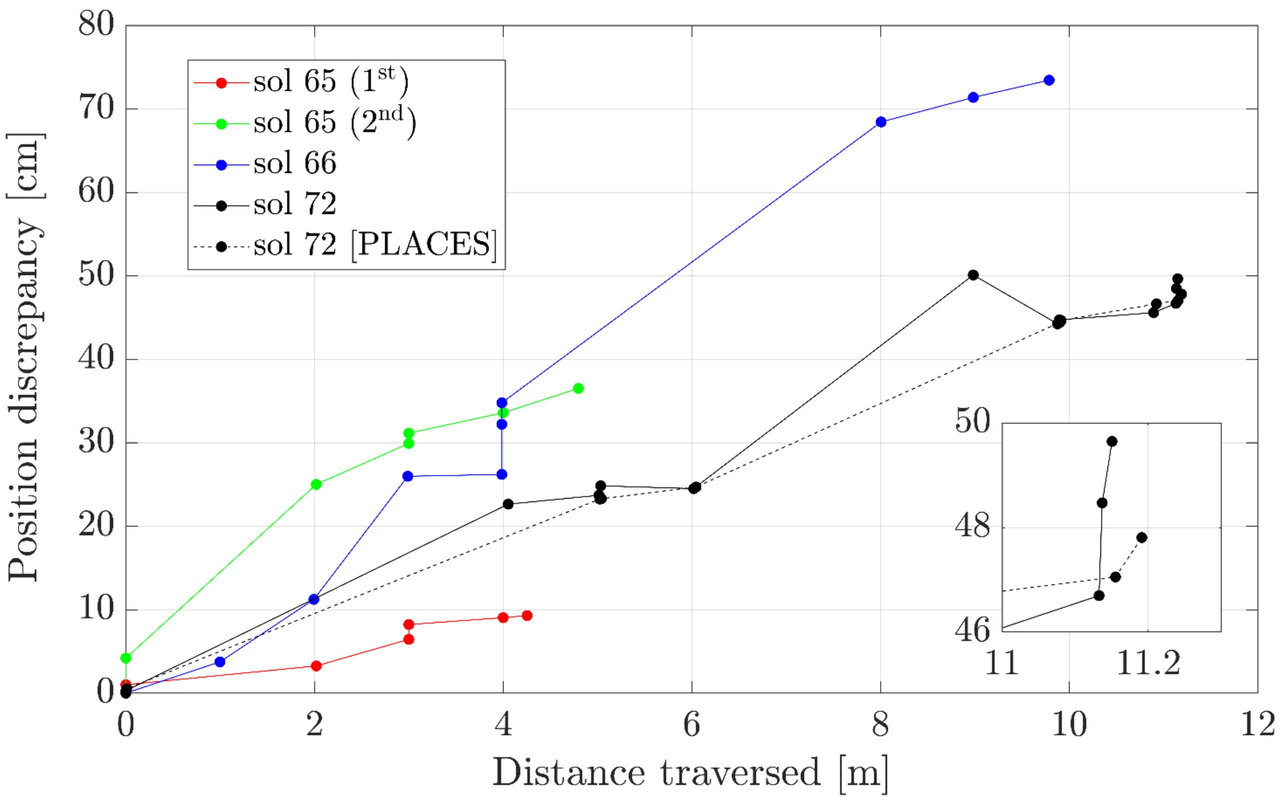

Figure 7. Position discrepancies between Perseverance's telemetry-based path and our VO-estimated trajectory. Telemetered rover's positions are retrieved from the image metadata (**solid**) and from the PLACES database (**dashed**), if pose updates were produced onboard. The points on the curves represent new stereo pair acquisitions, and traversed distances are based on telemetry data. Our estimates of short drive steps (<1 m) are consistent with telemetry-based paths, and larger discrepancies (10-30 cm) are detected for stereo pairs acquired more than 1 m apart. On sol 72, our VO-estimated path is more consistent with the refined pose estimated onboard (**black, dashed**), as shown in the bottom right panel.

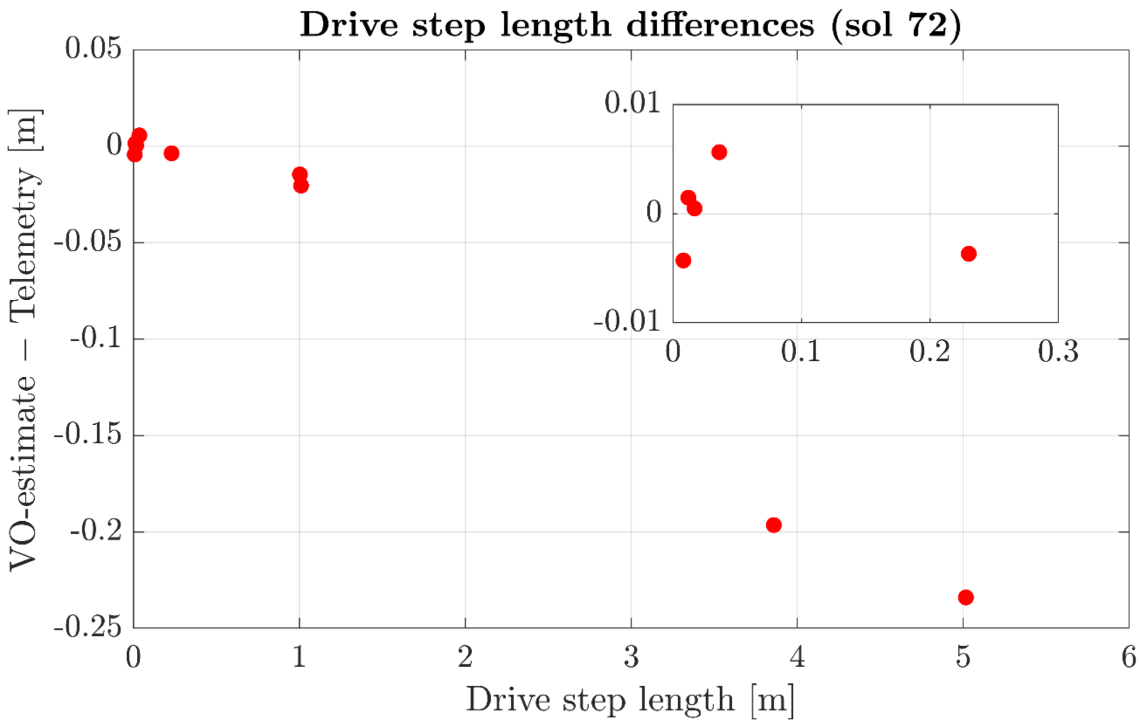

Figure 8. Differences between the distance travelled at each drive step on sol 72 accordingly to our VO estimate and to the refined localization solution produced onboard (retrieved from the PLACES database). The drive step length (*i.e.*, the distance traversed between the acquisition of adjacent stereo pairs), reported on the horizontal axis, is computed accordingly to the refined telemetry data. The two solutions are fully consistent (differences of 1-2 cm) for short distances (<1 m), leading to discrepancies <5 mm for drive steps <30 cm (top right panel). Larger discrepancies (~20 cm) are detected for drive steps >3.5 meters.

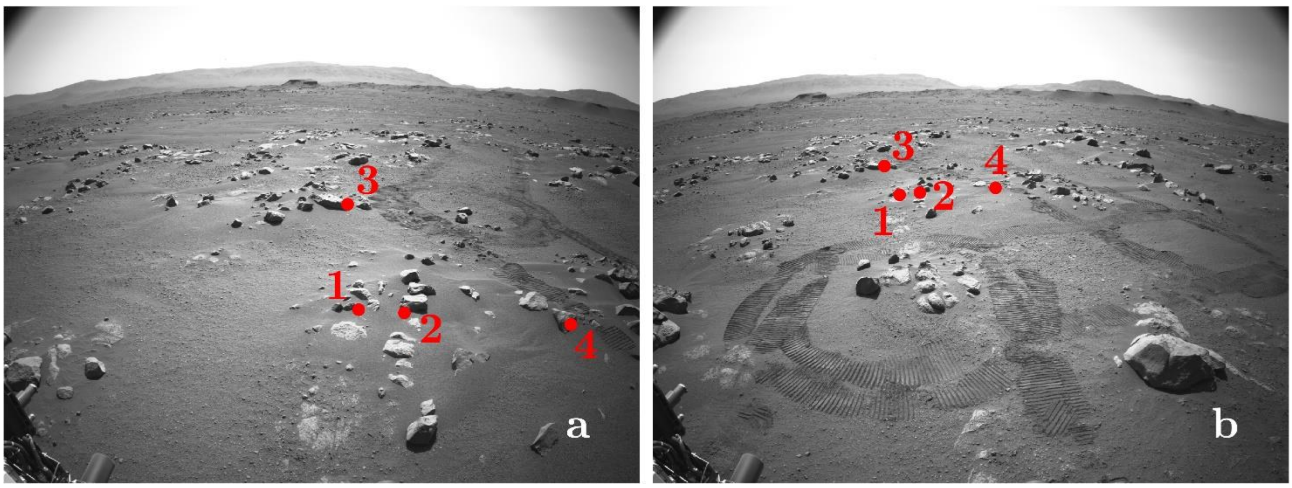

Figure 9. Left images acquired (**a**) at the beginning and (**b**) at the end of the first leg of the path driven on sol 65. Corner-points (**red**) corresponding to the same landmarks in the two images are labelled using same numbers.

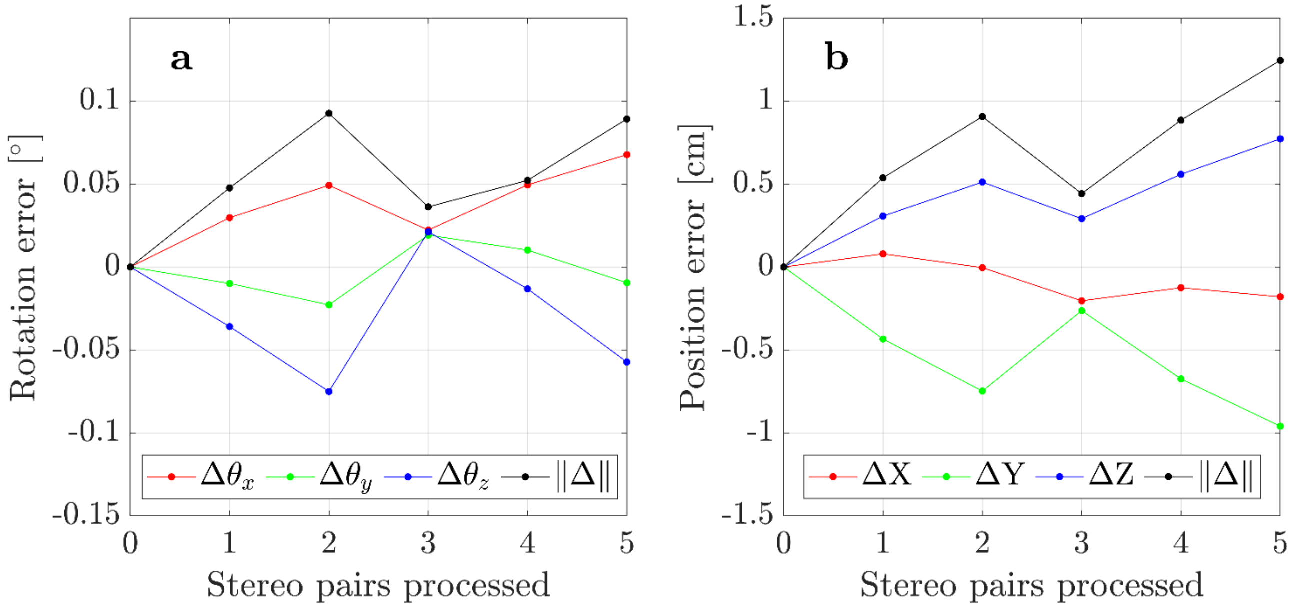

Figure 10. Attitude (**a**) and position (**b**) errors on sol 120. The total rotation error (**a, black**) is computed as the norm of the rotation error vector $\Delta = [\Delta\theta_x \quad \Delta\theta_y \quad \Delta\theta_z]$. The total position error (**b, black**) is computed as the norm of the position error vector $\Delta = [\Delta X \quad \Delta Y \quad \Delta Z]$. The overall position and attitude errors are <1.5 cm and <0.1°, respectively.