

# Rovers Localization by using 3D-to-3D and 3D-to-2D Visual Odometry

Simone Andolfo

Department of Mechanical and  
Aerospace Engineering  
Sapienza—University of Rome  
Rome, Italy

andolfo.1693615@Studenti.uniroma1.it

Flavio Petricca

Department of Mechanical and  
Aerospace Engineering  
Sapienza—University of Rome  
Rome, Italy

flavio.petricca@uniroma1.it

Antonio Genova

Department of Mechanical and  
Aerospace Engineering  
Sapienza—University of Rome  
Rome, Italy

antonio.genova@uniroma1.it

**Abstract**—Space robotic systems have been playing a crucial role in planetary exploration missions, expanding our access to harsh and hostile environments in the Solar System. Rovers' activities are still mainly controlled through ground operations, and our goal is to develop autonomous systems for navigation and path planning. The position estimates obtained by processing Wheel Odometry (WO) data induce significant errors because of wheels' loss of traction that is caused by, for example, high-slippage terrains (e.g., sandy-loose soils, steep slopes). Our work is focused on the implementation of a localization software based on Visual Odometry (VO). This is a technique developed for the estimation of rovers' position and attitude by using stereo images captured during the vehicle's motion. To determine the attainable accuracy of our software, we carried out a set of numerical simulations through a digitally-reproduced Martian-like environment. The results show that the algorithm allows reconstructing the rover's trajectory with higher accuracies compared to the localization system requirements of the NASA Mars Exploration Rovers (i.e., 10% error over a 100-m traverse [1]).

**Keywords**—Space Robotic Systems, Visual Odometry, Rovers

## I. INTRODUCTION

In the next decades, rovers will be designed to operate in harsh environments on the Martian surface (e.g., polar layered deposits). To enhance the scientific return of these missions, an accurate localization of the wheeled vehicle will be accomplished with different techniques. A precise knowledge of the rover's position and orientation (*pose*) across the target site enables a better planning of the future path and science investigations.

A basic method to update the rover's pose is to combine the attitude measurements provided by the onboard Inertial Measurement Units (IMUs) with a Wheel Odometry (WO)-based estimate of the distance travelled by the rover. However, when the rover moves across high-slippage terrains, the wheels may lose traction, leading to lower accuracies of the WO measurements that inflate errors in the position update. Visual Odometry (VO) is then integrated with WO+IMU to significantly enhance the trajectory reconstruction of planetary rovers in such critical scenarios.

VO is the technique that sequentially estimates the pose of a robotic vehicle by processing the images captured during the motion. VO can be exploited using the following camera configurations: (1) stereo cameras [1-2], (2) monocular cameras [3], (3) omnidirectional cameras [4] and (4) multi-camera systems [5]. The images also provide information regarding the site map, and Simultaneous Localization and Mapping (SLAM) problems are implemented to combine these observations with other datasets as, for example, LIDAR range measurements [6].

NASA missions Mars Exploration Rovers (MER) [1-2] and Mars Science Laboratory (MSL) [7-8] successfully used *stereo* VO to precisely update the rover's position during science operations on the Martian surface. VO currently represents a standard method for planetary exploration missions, including NASA Perseverance rover (Mars2020 mission) [9], Chinese lunar rover Yutu-2 [10], ESA-Roscosmos ExoMars rover [11-12].

Stereo VO algorithms can be distinguished into two main categories, that is 3D-to-3D and 3D-to-2D methods [13]. The first class of algorithms is based on the estimation of the rover's pose as a rigid transformation through the alignment of the 3D point-clouds computed before and after the motion step. The 3D-to-2D methods provide an update of the rover location and orientation by minimizing the distance between the reprojection of the 3D point cloud obtained through landmarks triangulation from the previous stereo pair onto one of the new images acquired after the motion step and the *real* corresponding 2D points (*reprojection error*).

In this paper we present numerical simulations of the rover navigation based on a 3D-to-3D and a 3D-to-2D VO-based localization software. Synthetic images are generated and processed to assess the trajectory reconstruction accuracies in a Martian-like scenario. After introducing the motion parameters, we describe the motion estimation algorithms adopted in this study. These methods are used for a 100-m rover's traverse to compare their performances.

## II. PARAMETRIZATION OF THE ROVER'S MOTION

To estimate the 6 degrees of freedom motion step, the stereo VO algorithm takes as input two pairs of stereo images acquired before and after the motion step and a first-guess motion estimate based on WO and IMU measurements.

Let  $\{B\}$  and  $\{A\}$  denote the right-handed coordinate systems attached to the stereo camera before and after the rover motion step. The rigid body transformation that converts  $\{B\}$  into  $\{A\}$  consists of a rotation defined by  $\mathbf{R}$  (i.e., the  $(3 \times 3)$  rotation matrix from  $\{B\}$  to  $\{A\}$ ) and  $\boldsymbol{\tau}$  (i.e., the  $(3 \times 1)$  position vector from  $\mathbf{O}_A$  to  $\mathbf{O}_B$  whose components are expressed in  $\{A\}$ ). In this work, the rotation is expressed by means of the rotation vector  $\boldsymbol{\theta} = [\theta_x, \theta_y, \theta_z]^T$  that collects the triplet of angles associated with the (3-2-1) sequence of elementary rotations (i.e., Bryant angles).

The subscripts "L" and "R" will be used to refer to quantities associated with left or right images. A pinhole camera model is assumed for the onboard cameras, which have parallel optical axes and acquire *rectified* images (i.e., their corresponding epipolar lines coincide and are parallel to the x-axis of image) that are not affected by distortion effects.

Given a physical point  $\mathcal{P}$  in the 3D space (i.e., a landmark), its projections onto left and right image planes are

two image-points, denoted by  $\mathbf{p}_L$  and  $\mathbf{p}_R$ . By means of stereo triangulation, the 3D coordinates of  $\mathcal{P}$  can be estimated through the processing of 2D coordinates of the associated image features. The resulting (3×1) vector is defined with respect to the “actual” stereo camera frame (*i.e.*, {B} or {A}) depending on which stereo pair is considered) (Fig. 1). If a landmark  $\mathcal{P}$  is observed both before and after the rover’s motion, it is possible to compute both the vectors  $\mathbf{P}^{(B)}$  and  $\mathbf{P}^{(A)}$ . They are related by the motion equation:

$$\mathbf{P}^{(A)} = \mathbf{R}(\boldsymbol{\theta})\mathbf{P}^{(B)} + \boldsymbol{\tau} \Leftrightarrow \begin{bmatrix} \mathbf{P}^{(A)} \\ 1 \end{bmatrix} = \mathbf{T} \begin{bmatrix} \mathbf{P}^{(B)} \\ 1 \end{bmatrix} \quad (1)$$

where  $\mathbf{T}$  is the (4×4) transformation matrix  $\mathbf{T}$  from {B} to {A} defined as:

$$\mathbf{T} = \begin{bmatrix} \mathbf{R} & \boldsymbol{\tau} \\ \mathbf{0}_{1 \times 3} & 1 \end{bmatrix} \quad (2)$$

### III. METHODS

#### A. Features detection

The first step of VO algorithms is the detection of features in a new stereo pair. The images are processed by a corner detector (*e.g.*, the Förstner interest operator [14]) that extracts a set of possible trackable features (Fig. 2). To identify corner-points, the corner detector measures the image gradient (*i.e.*, the variation of the brightness intensity of the pixels) inside a small image patch defined by a pixel with coordinates (x,y), and computes a “corner strength” function  $f(x,y)$  [15]. Pixels characterized by a high corner strength are more likely to be corner-points. Moreover, subpixel interpolation techniques are used to better localize a feature. Once a local-maximum of  $f(x,y)$  is detected, the first-guess position of the corner is refined to sub-pixel accuracy by fitting a 2D quadratic function to  $f(x,y)$  evaluated in the local 3×3 neighbourhood and finding its maximum [15]. This operation also yields a (2×2) covariance matrix  $\boldsymbol{\Sigma}_{\hat{\mathbf{p}}}$  associated with the improved corner coordinates  $\hat{\mathbf{p}}$ , which are perturbed by a white Gaussian error vector  $\mathbf{e}_{\hat{\mathbf{p}}}$  with covariance  $\boldsymbol{\Sigma}_{\hat{\mathbf{p}}}$ .

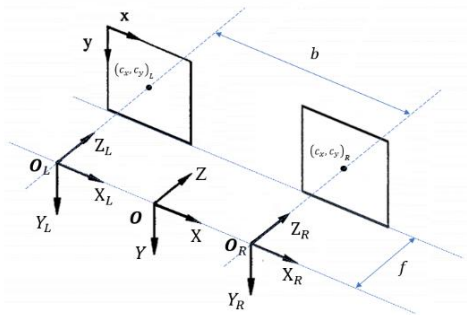


Fig. 1. Stereo camera frame ( $\mathcal{O}$ , X, Y, Z), stereo baseline and focal length.

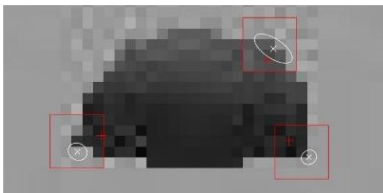


Fig. 2. Detail of three detected corner-points. For each of them, the white cross marks the best estimated corner-point, the ellipse is the 1- $\sigma$  covariance ellipse (associated with  $\boldsymbol{\Sigma}_{\hat{\mathbf{p}}}$ ), and the red square is the optimal cross-correlation window for the stereo matching.

Image noise can cause erroneous detections of outliers as corner-points. To prevent from false detections, outlier rejection techniques are adopted in the stereo matching step.

#### B. First Stereo Matching

Once corner-points are extracted from the first stereo pair, a stereo match is obtained by pairing a corner in the left image  $\hat{\mathbf{p}}_L$  with the corresponding corner  $\hat{\mathbf{p}}_R$  in the right image.

The stereo matching technique adopted in this study is the *block matching*. To establish if we have a match between two corners, square windows are defined to outline their location, and the Normalized Cross-Correlation (NCC) cost index is computed [16]. The NCC quantifies the grade of similarity between the image portions framed by the left and the right windows. It is defined as follows:

$$NCC = \frac{\sum_{k=1}^M [(I_L(k) - \bar{I}_L)(I_R(k) - \bar{I}_R)]}{\sqrt{\sum_{k=1}^M [(I_L(k) - \bar{I}_L)^2 (I_R(k) - \bar{I}_R)^2]}} \quad (3)$$

where  $I_m(k)$  denotes the brightness intensity of the  $k$ -th pixel inside the  $m=L,R$  correlation window, while  $\bar{I}_m(k)$  denotes the mean intensity of the  $M$  pixels inside the  $m$  window. The side of the windows should be an odd number of pixels, so that a central pixel can be identified without any ambiguity.

However, the adoption of the NCC as the only matching criterion leads to many mismatches, which can be discarded by imposing the epipolar and the uniqueness constraints. According to the epipolar constraint, two corresponding features should be approximately aligned (*i.e.*, the epipolar), if the cameras are calibrated and, consequently, the images rectified. Therefore, in case the y-coordinates of two highly correlated corners significantly differ, *i.e.*,  $|y_L - y_R| > \Delta y$ , a mismatch is detected. We assumed a  $\Delta y$  threshold of 1 pixel. The uniqueness constraint states that a feature in one image has at most one corresponding feature in the other image. In case a left feature fulfils the epipolar constraint with multiple right features, we select the matching with the highest value of the NCC index.

After a successful feature matching, stereo triangulation is used to estimate the 3D coordinates of the associated landmark with respect to the stereo camera frame. Let  $(x_L, y_L)$  and  $(x_R, y_R)$  denote the coordinates of two matched features and  $\boldsymbol{\Sigma}_{\hat{\mathbf{p}}}_L$  and  $\boldsymbol{\Sigma}_{\hat{\mathbf{p}}}_R$  denote the associated (2×2) covariance matrices. The estimated landmark’s coordinates  $\hat{\mathbf{P}} = [\hat{X}, \hat{Y}, \hat{Z}]^T$  are obtained according to:

$$\hat{X} = b (x_L + x_R - 2c_x) / [2(x_L - x_R)] \quad (4)$$

$$\hat{Y} = [b s_x (y_L + y_R - 2c_y)] / [2 s_y (x_L - x_R)] \quad (5)$$

$$\hat{Z} = b s_x / (x_L - x_R) \quad (6)$$

where  $b$  is the stereo baseline;  $s_x$  and  $s_y$  are the focal length  $f$  (*i.e.*, the distance between the lens and the image plane) scaled with horizontal and vertical pixel dimension, respectively; and  $c_x$  and  $c_y$  are the image-coordinates of the central image pixel.

The error of the triangulated landmark position  $\mathbf{e}_{\hat{\mathbf{p}}}$  is non-Gaussian because of the non-linear equations eq. (4)-(6). However, we used standard propagation methods to approximate  $\mathbf{e}_{\hat{\mathbf{p}}}$  as Gaussian noise with zero mean and covariance  $\boldsymbol{\Sigma}_{\hat{\mathbf{p}}} = \mathbf{J} \boldsymbol{\Sigma} \mathbf{J}^T$ , where the Jacobian  $\mathbf{J}$  is a (3×4) matrix

with the partial derivatives of the 3D landmark coordinates with respect to  $x_L, y_L, x_R, y_R$ , and  $\Sigma = \text{diag}(\Sigma_p^L, \Sigma_p^R)$  [17].

### C. Tracking

At the end of the motion step, a second stereo pair is acquired and processed by the corner detector. The tracking of the previously stereo-matched image-features is then necessary to identify the corresponding corner-points in the new pair (Fig. 3). An accurate tracking is crucial to avoid erroneous matching that would significantly affect the quality of the motion estimate.

To accomplish this task, the 3D points  $\hat{\mathbf{P}}_i^{(B)}$  are updated according to a first-guess motion prediction based on WO and IMU measurements. We assumed white Gaussian noises for WO and IMU with standard deviations of 5 cm and  $0.05^\circ$ , respectively.

The vectors based on WO and IMU predictions are then reprojected onto the left image of the second stereo pair to obtain a set of 2D image-points ( $\hat{\mathbf{p}}_{L,i}^{(A)}$ ). The tracking step consists of:

- detection of corners  $\hat{\mathbf{p}}_{L,k}^{(A)}$  within a  $30 \times 30$  pixels subimage portion for each predicted corner  $\hat{\mathbf{p}}_{L,i}^{(A)}$ ;
- computation of the NCC index for each pair  $\hat{\mathbf{p}}_{L,i}^{(B)}$  and  $\hat{\mathbf{p}}_{L,k}^{(A)}$ ;
- selection of the pair with the highest  $\text{NCC} > 5$ ; if all NCC indexes are lower than 5 no match is found.

### D. Motion Estimate

#### 1) 3D-to-3D methods

The 3D-to-3D methods require a second stereo matching to track the left image corners in the new right image. The 3D coordinates of the landmarks are then triangulated and their covariances are computed. These quantities are respectively denoted by  $\hat{\mathbf{P}}_i^{(A)}$  and  $\Sigma_i^{(A)}$ , with  $i = 1, \dots, N$ .

This method is based on a maximum-likelihood estimation (MLE) of the motion parameters. By processing the triangulated coordinates of the same set of landmarks before,  $\hat{\mathbf{P}}_i^{(B)}$ , and after,  $\hat{\mathbf{P}}_i^{(A)}$ , the motion, the filter provides an estimate of the rover's pose by minimizing the distance between the two point clouds. The solution is computed including information on the 3D distribution of the triangulation error through the covariance matrices  $\Sigma_i^{(B)}$  and  $\Sigma_i^{(A)}$ .

The cost function to be minimized depends on the residual errors  $\mathbf{e}_i = \hat{\mathbf{P}}_i^{(A)} - \mathbf{R} \hat{\mathbf{P}}_i^{(B)} - \boldsymbol{\tau}$  and is defined as follows [17]:

$$E_{3D} = \sum_i (\mathbf{e}_i^T \mathbf{W}_i \mathbf{e}_i) \quad (7)$$

A first guess of the rover's pose is obtained by minimizing this function through a least-squares estimator (LSE). The resulting rotation matrix and translation vector are used as *a priori* solution for the MLE that enables a refined estimate of these motion parameters. The MLE algorithm enables a different weighting of each component of the error vector  $\mathbf{e}_i$  by accounting for a full covariance matrix that is obtained from the full covariance matrices associated with the triangulated landmarks coordinates. In the preliminary LSE solution the weighting factors are given by diagonal matrices

$\mathbf{W}_i = w_i \mathbf{I}$  that account for a uniform observables accuracy [17].

The weights used for the MLE approach are defined as:

$$\mathbf{W}_i = (\Sigma_i^{(A)} + \mathbf{R}_0 \Sigma_i^{(B)} \mathbf{R}_0^T)^{-1} \quad (8)$$

which is the inverse of the covariance matrix associated with the *linearized* residual errors, obtained by linearizing the motion equations eq. (1) about a first-guess solution for the rover's motion step, *i.e.*,  $\mathbf{R}_0, \boldsymbol{\tau}_0$ .

The MLE is iterated until convergence that is declared when the criterion  $|\boldsymbol{\theta}_k - \boldsymbol{\theta}_{k-1}| < \sigma$  is met (the subscript  $k$  denotes the current iteration).

To compute an optimal motion estimate, a random sample consensus (RANSAC)-like process is embedded in the algorithm. A number  $N_{\text{sol}}$  of maximum-likelihood motion estimates are computed, and the best solution allows minimizing the cost index defined as:

$$S_k = \sum_{i=1}^N (\|\hat{\mathbf{p}}_{L,i}^{(A)} - \bar{\mathbf{p}}_{L,i,k}^{(A)}\|^2 + \|\hat{\mathbf{p}}_{R,i}^{(A)} - \bar{\mathbf{p}}_{R,i,k}^{(A)}\|^2) \quad (9)$$

where  $\hat{\mathbf{p}}_{m,i}^{(A)}$  is the corner-point detected in the new (*i.e.*, acquired after the motion step)  $m = L, R$  image and corresponding to landmark  $i$ , while  $\bar{\mathbf{p}}_{m,i,k}^{(A)}$  is the reprojection (onto the same image) of the 3D point obtained updating  $\hat{\mathbf{P}}_i^{(B)}$  according to the  $k$ -th solution for the motion parameters, with  $k = 1, \dots, N_{\text{sol}}$ . Each solution is obtained by processing the observables associated with  $n$  landmarks randomly chosen among the  $N$  successfully tracked ones ( $n$  can range from 8-12 up to  $N$ ). By computing multiple solutions, the less accurate ones are discarded (*e.g.*, solutions relying on noisy features or outliers that have not been rejected).

#### 2) 3D-to-2D motion estimate

Another approach to estimate the rover's motion is based on the 3D-to-2D formulation, which does not require the second stereo triangulation step. The optimal motion parameters  $\mathbf{R}$  and  $\boldsymbol{\tau}$  are those that minimize the reprojection error, that is the cost function defined as:

$$E_{2D} = \sum_{i=1}^N \|\hat{\mathbf{p}}_{h,i}^{(A)} - \bar{\mathbf{p}}_{h,i}^{(A)}(\hat{\mathbf{P}}_i^{(B)}, \mathbf{R}, \boldsymbol{\tau})\|^2 \quad (10)$$

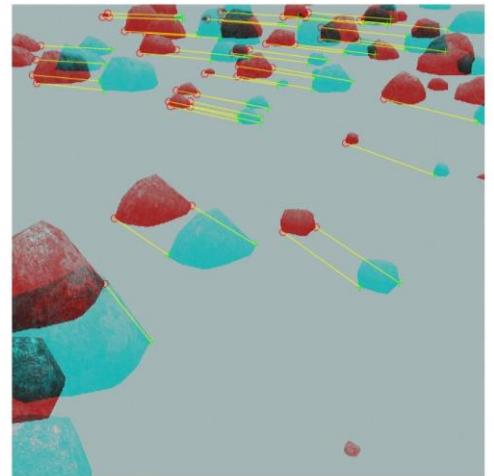


Fig. 3. Rototranslational motion step. Left images of the first and the second stereo pairs are shown with reddish and greenish-blue colours, respectively. Tracked features are linked with yellow lines.

where  $\hat{p}_{h,i}^{(A)}$  is the corner-point detected in the new (*i.e.*, acquired after the motion step)  $h = L, R$  image and corresponding to landmark  $i$ , and  $\hat{p}_{h,i}^{(A)}$  is the reprojection (onto the same image) of the 3D point obtained updating  $\hat{P}_i^{(B)}$  according to  $\mathbf{R}$  and  $\boldsymbol{\tau}$ .

Eq. (10) only takes into account two consecutive pairs of stereo images, which are captured at times  $t_k$  and  $t_{k-1}$ . The localization system of the ExoMars rover will rely on a 3D-to-2D formulation for the motion estimation problem with a modified cost function that includes the observations of the same landmark in the  $n_i$  previously acquired images [12]. To mitigate this growing error, a local bundle adjustment (BA) is carried out when the rover completes a path planned step [12]. This approach enables a joint refinement of the estimated rover's motion parameters and the 3D landmarks coordinates.

To compare the achievable accuracies of rover's pose estimation with 3D-to-3D and 2D-to-2D methods, we implemented both motion estimate algorithms without the adjustment of the landmark locations. We assumed then that the landmarks coordinates are fixed and we estimated the camera ego-motion only (motion-only bundle adjustment [18]). Furthermore, we considered the 2D image points from the left camera only.

The motion estimate procedure consists of two steps. A first-guess motion hypothesis  $\mathbf{R}_0$  and  $\boldsymbol{\tau}_0$  is obtained solving the perspective-three-points (P3P) problem [19]. A non-linear optimization process is then applied to refine the solution. To remove outliers, a RANSAC process is embedded in the computation of the motion hypothesis.  $N_{sol}$  solutions are computed solving the P3P problem by using sets of 3D-2D correspondences randomly selected. The best solution is retrieved by minimizing the global reprojection error. A final non-linear optimization is carried out using the Levenberg-Marquardt (LM) algorithm, which is used to solve local and global BA problems, involving multiple camera views and thousands of points [20].

The 2D reprojection point  $\hat{p}_{L,i}^{(A)} = [\bar{x}_L, \bar{y}_L]_i^T$  associated with landmark  $i$  can be computed according to:

$$\begin{bmatrix} \hat{p}_{L,i}^{(A)} \\ 1 \end{bmatrix} = \frac{1}{Z} \mathbf{K} \begin{bmatrix} \bar{P}_i^{(A_L)} \\ 1 \end{bmatrix} = \frac{1}{Z} \mathbf{K} \mathbf{T} \begin{bmatrix} \bar{P}_i^{(B_L)} \\ 1 \end{bmatrix} \quad (11)$$

where  $\mathbf{K}$  is the intrinsic camera parameters matrix [13];  $\mathbf{T}$  is the transformation matrix defined by  $\mathbf{R}$  and  $\boldsymbol{\tau}$  (see eq. (2));  $\bar{P}_i^{(B_L)} = [X', Y', Z']^T$  are the triangulated coordinates of landmark  $i$  before the rover's motion;  $\bar{P}_i^{(A_L)} = [X, Y, Z]^T$  are the updated coordinates of landmark  $i$  computed according to  $\mathbf{R}$  and  $\boldsymbol{\tau}$ . In eq. (11), the 3D points  $\bar{P}_i^{(B_L)}$  and  $\bar{P}_i^{(A_L)}$  are defined in the *left* camera frames before and after the rover's motion. Similarly, the estimated motion parameters  $\mathbf{R}$  and  $\boldsymbol{\tau}$  link two successive poses of the *left* camera frames.

To evaluate the Jacobian matrix  $\mathbf{J}$ , we adopt an exponential parametrization to represent the motion parameters [21]. The new set of motion variables are defined by the two (3×1) vectors  $\mathbf{t}$  and  $\boldsymbol{\omega}$ , related to  $\boldsymbol{\tau}$  and  $\mathbf{R}$  as follows:

$$\mathbf{T} = \begin{bmatrix} e^{\tilde{\boldsymbol{\omega}}} & \mathbf{V}\mathbf{t} \\ \mathbf{0}_{1 \times 3} & 1 \end{bmatrix} = \begin{bmatrix} \mathbf{R} & \boldsymbol{\tau} \\ \mathbf{0}_{1 \times 3} & 1 \end{bmatrix} \quad (12)$$

where:

$$e^{\tilde{\boldsymbol{\omega}}} = \mathbf{I} + \frac{\sin(\alpha)}{\alpha} \tilde{\boldsymbol{\omega}} + \frac{1 - \cos(\alpha)}{\alpha^2} \tilde{\boldsymbol{\omega}}^2 \quad (13)$$

$$\mathbf{V} = \mathbf{I} + \frac{1 - \cos(\alpha)}{\alpha^2} \tilde{\boldsymbol{\omega}} + \frac{\alpha - \sin(\alpha)}{\alpha^3} \tilde{\boldsymbol{\omega}}^2 \quad (14)$$

with  $\alpha = |\boldsymbol{\omega}|$  and  $\tilde{\boldsymbol{\omega}}$  being the skew-symmetric matrix generated by the vector  $\boldsymbol{\omega}$  [21]. The computation of the Jacobian matrix is required to solve the LM augmented normal equations, which provide the correction  $\delta\mathbf{v}$  for the *a priori* motion parameters  $\mathbf{v}_0 = [\mathbf{t}_0^T, \boldsymbol{\omega}_0^T]^T$  as follows:

$$(\mathbf{J}^T \mathbf{J} + \lambda \mathbf{D}) \delta\mathbf{v} = \mathbf{J}^T \boldsymbol{\epsilon} \quad (15)$$

From the same motion hypothesis  $\mathbf{v}_0$ , multiple refined solutions are then obtained by randomly selecting and processing  $N=8$  2D-3D correspondences; the best solution is finally chosen as the one that minimizes the global reprojection error.

#### IV. NUMERICAL SIMULATIONS

Synthetic images of a simulated rough Martian terrain were generated to run numerical simulations of the planetary rover traverse. The site on the Martian surface was created by defining randomly distributed hazards along the path of the rover. The trajectory and attitude of the rover were simulated by using sequential poses of the stereo camera frame relative to the inertial frame {I}. The position of the rover is based on a simplified kinematic model, which is well-suited to trajectory estimators based on the analysis of imaging data.

In our numerical simulations, we assumed that a new pair of stereo images was acquired with a 1-s sampling rate. The cameras were modelled with a fixed downward tilt of 30° and the properties of the Navigation Cameras (NavCams) of the MER rovers [22] (Table 1).

#### V. RESULTS

A 100-m traverse is simulated, and both the 3D-to-3D and the 3D-to-2D localization algorithms are used to reconstruct the rover's trajectory. We defined waypoints to enable a rectilinear trajectory parallel to the Y-axis of the frame {I}. The distance  $D$  between two successive waypoints has been

TABLE I. MAIN MER NAVCAM PROPERTIES [22]

Parameter	Image size	Pixel dimensions	Focal length	Stereo baseline	$s_x, s_y$	$c_x, c_y$
Value	1024×1024 pixels	12×12 μm	14.67 mm	20.0 cm	1222.5	511.5 pixels

set equal to 50 cm, which is in line with the camera time step of image acquisition (1 s).

After processing each pair of stereo images, the final trajectory is retrieved combining every motion step. Given the motion parameters defining the  $k$ -th motion step (*i.e.*,  $\mathbf{R}_{k-1}^k$  and  $\boldsymbol{\tau}^{(k)}$ ), the position of the rover is updated according to:

$$\mathbf{P}_k^{(I)} = \mathbf{P}_{k-1}^{(I)} + \mathbf{R}_{k-1}^{(I)} (\mathbf{R}_{k-1}^k)^T (-\boldsymbol{\tau}^{(k)}) \quad (16)$$

while the attitude angles  $\boldsymbol{\theta}_k$  defining the actual camera orientation with respect to  $\{I\}$  are extracted from the rotation matrix:

$$\mathbf{R}_{\{I\}}^k = \mathbf{R}_{k-1}^k (\mathbf{R}_{k-1}^{(I)})^T \quad (17)$$

The variables  $\boldsymbol{\theta}_k$  and  $\mathbf{P}_k^{(I)}$  are collected by the vector  $\mathbf{C}_k$  that defines the actual camera pose with respect to  $\{I\}$ .

Fig. 4 shows a comparison of the position error obtained using the localization methods described in this work as a function of the distance travelled by the rover. Both VO techniques provides better accuracies compared to WO+IMU only, which lead to a final position error of  $\sim 8.3$  m. The more stable trends of the position errors with both 3D-to-3D and 3D-to-2D VO methods yield accuracies of better than 15 cm and 25 cm, respectively.

The uncertainties on the actual camera pose  $\mathbf{C}_k$  are computed from the associated covariance matrix  $\boldsymbol{\Sigma}_k$  that is obtained using standard error propagation methods, which approximates  $\mathbf{C}_k$  as a random variable affected by white Gaussian noise with covariance  $\boldsymbol{\Sigma}_k$ . However, due to the high nonlinearity of the problem, a refined uncertainty analysis would require Monte Carlo simulations [23].

Given the  $(6 \times 6)$  covariance matrix  $\boldsymbol{\Sigma}_{k-1}$  associated with the  $(k-1)$ -th pose  $\mathbf{C}_{k-1}$  and the  $(6 \times 6)$  covariance matrix  $\boldsymbol{\Sigma}_{k-1}^k$  associated with the estimated motion step, the covariance  $\boldsymbol{\Sigma}_k$  associated with the  $k$ -th pose  $\mathbf{C}_k$  can be computed as:

$$\boldsymbol{\Sigma}_k = \mathbf{J}_k \begin{bmatrix} \boldsymbol{\Sigma}_{k-1} & \mathbf{0}_{6 \times 6} \\ \mathbf{0}_{6 \times 6} & \boldsymbol{\Sigma}_{k-1}^k \end{bmatrix} \mathbf{J}_k^T \quad (18)$$

Let  $\mathbf{M}_{k-1}^k$  denote the  $(6 \times 1)$  vector collecting the estimated motion parameters  $\boldsymbol{\theta}_{k-1}^k$  and  $\boldsymbol{\tau}^{(k)}$  defining the motion step from  $\mathbf{C}_{k-1}$  to  $\mathbf{C}_k$ . Then, it results that:

$$\mathbf{J}_k = \begin{bmatrix} \frac{\partial \mathbf{C}_k}{\partial \mathbf{C}_{k-1}} & \frac{\partial \mathbf{C}_k}{\partial \mathbf{M}_{k-1}^k} \end{bmatrix} \quad (19)$$

Fig. 5 shows the  $3\text{-}\sigma$  formal uncertainties of the pose vector that are computed from the covariance matrix  $\boldsymbol{\Sigma}_{k-1}^k$  resulting from the MLE solution, supporting a full consistency with the estimation errors. An a priori uncertainty of 5 mm and  $0.05^\circ$  is assumed for each component of the position and attitude vectors, respectively.

## VI. SUMMARY AND FUTURE WORK

The main goal of this work is the implementation of a stereo VO algorithm for precise rover localization.

A detailed description of the algorithms is provided, including the estimation scheme based on 3D-to-3D and 3D-to-2D methods.

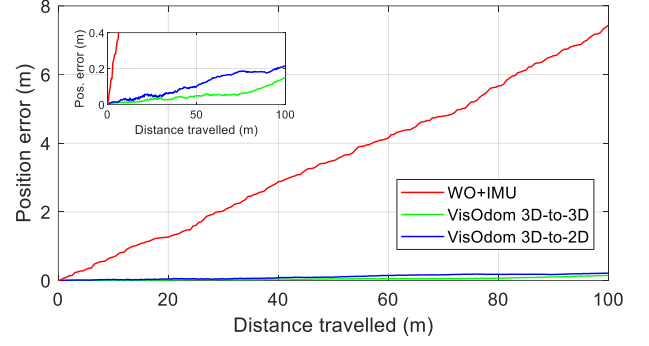


Fig. 4. 100 meters traverse: position error.

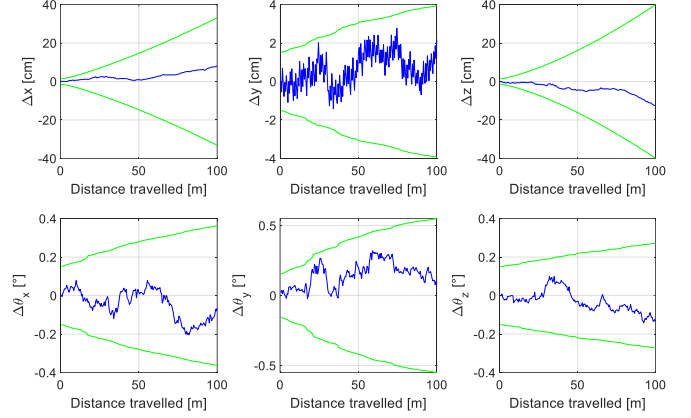


Fig. 5. Position and attitude error vectors. The mean value of the error is depicted in blue, while the green curves are the  $3\text{-}\sigma$  formal uncertainties.

A mission scenario that accounts for a 100-m traverse is then presented to assess the attainable accuracy of the implemented VO methods. The results prove that both VO algorithms significantly enhance WO+IMU estimation, leading to an accurate reconstruction of the rover's position and attitude.

A future development of our work will consist in the integration of these trajectory estimators in a Guidance Navigation and Control system for planetary rovers. The methods presented in this study will be extended by including the mapping of the environment and an accurate characterization of the terrain properties.

## ACKNOWLEDGEMENTS

Our software is based on Python and Matlab subroutines that interface with Gazebo simulation environment [24].

## REFERENCES

- [1] M. Maimone, Y. Cheng, and L. Matthies, "Two Years of Visual Odometry on the Mars Exploration Rovers", *J. Field Robot.*, vol. 24(3), pp. 169–186, 2007.
- [2] L. Matthies, M. Maimone, A. Johnson, et al., "Computer Vision on Mars", *Int. J. Comput. Vis.*, vol. 75, pp. 67–92, 2007.
- [3] D. Nister, O. Naroditsky and J. Bergen, "Visual odometry", *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004 (CVPR 2004)*.
- [4] Z. Zhang, H. Rebecq, C. Forster, and D. Scaramuzza, "Benefit of Large Field-of-View Cameras for Visual Odometry". In: *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)* (2016), pp. 801–808.
- [5] C. Forster, Z. Zhang, M. Gassner, et al., "SVO: Semidirect Visual Odometry for Monocular and Multicamera Systems", in *IEEE Transactions on Robotics*, vol. 33, no. 2, pp. 249–265, April 2017.

- [6] C. Debeunne, and D. Vivet, "A Review of Visual-LiDAR Fusion based Simultaneous Localization and Mapping", *Sensors*, vol. 20, no. 7, 2020.
- [7] J. P. Grotzinger, J. Crisp, A. R. Vasavada, et al., "Mars Science Laboratory Mission and Science Investigation", *Space Sci. Rev.*, vol. 170(1–4), 2012.
- [8] A. Rankin, M. Maimone, J. Biesiadecki, et al., "Driving Curiosity: Mars Rover Mobility Trends During the First Seven Years", 2020 IEEE Aerospace Conference, pp. 1-19, 2020.
- [9] J. N. Maki, D. Gruel, and C. McKinney, "The Mars 2020 Engineering Cameras and Microphone on the Perseverance Rover: A Next-Generation Imaging System for Mars Exploration", *Space. Sci. Rev.*, vol. 216, 2020.
- [10] Y. Ma, S. Liu, B. Sima, et al., "A precise visual localisation method for the Chinese Chang'e-4 Yutu-2 rover". *The Photogrammetric Record*. 35(169), pp. 10-39, 2020.
- [11] M. Winter, C. Barclay, V. Pereira, et al., "ExoMars Rover Vehicle: Detailed Description of the GNC System". In: 13th Symposium on Advanced Space Technologies in Robotics and Automation, 2015 May 11–13; ESTEC, Noordwijk, Netherlands.
- [12] F. Souvannavong, C. Lemaréchal, L. Rastel, et al., "Vision-based motion estimation for the ExoMars rover", CNES (The French Space Agency), France, 2010.
- [13] D. Scaramuzza, and F. Fraundorfer, "Visual odometry Part I: the first 30 years and fundamentals", *IEEE Robot. Autom. Mag.* (2011) 80–92.
- [14] W. Förstner, and E. Gülch, "A fast operator for detection and precise location of distinct point, corners and centres of circular features". In: Proceedings of ISPRS intercommission conference on fast processing of photogrammetric data; 1987 Jun 2–4; Interlaken, Switzerland.
- [15] M. Brown, R. Szeliski, and S. Winder, "Multi-image matching using multi-scale oriented patches". In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR); 2005 June 20–25; San Diego, CA.
- [16] M. Z. Brown, D. Burschka, and G. D. Hager, "Advances in Computational Stereo", *IEEE Trans. Pattern. Anal. Mach. Intell.*, vol. 25, pp. 993–1008, 2003.
- [17] L. Matthies, and S.A. Shafer, "Error modeling in stereo navigation", *IEEE J. Robot. Autom.*, vol. 3, no. 3, pp. 239–248, 1987.
- [18] R. Mur-Artal, J. M. M. Montiel, and J. D. Tardós, "ORB-SLAM: A Versatile and Accurate Monocular SLAM System", in *IEEE Transactions on Robotics*, vol. 31, no. 5, pp. 1147-1163, Oct. 2015.
- [19] X. Gao, X. Hou, J. Tang, and H. Cheng, "Complete solution classification for the perspective-three-point problem", in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 8, pp. 930-943, Aug. 2003.
- [20] B. Triggs, P. Mclauchlan, R. Hartley, and A. Fitzgibbon. "Bundle Adjustment – A Modern Synthesis", in *International Workshop on Vision Algorithms*, pp. 298-372, 1999.
- [21] J. Blanco, "A tutorial on SE(3) transformation parameterizations and on-manifold optimization". University of Malaga, Tech. Rep, 2010.
- [22] J. N. Maki, J. F. Bell III, K. E. Herkenhoff, et al., "Mars Exploration Rover Engineering Cameras", *J. Geophys. Res.*, vol. 108(E12), 2003.
- [23] M. Pertile, S. Chiodini, S. Debei, and E. Lorenzini, "Uncertainty comparison of three visual odometry systems in different operative conditions", *Measurement*, pp. 388–396, vol. 78, 2016.
- [24] N. Koenig, and A. Howard, "Design and use paradigms for Gazebo, an open-source multi-robot simulator". In: *IEEE/RJS International Conference on Intelligent Robots and Systems (IROS)*; 2004 Sep 28–Oct 2; Sendai, Japan.