Scale and 2D Relative Pose Estimation of Two Rovers Using Monocular Cameras and Range Measurements

Chen Zhu[1], Gabriele Giorgi[1], and Christoph Günther[1,2]

[1] *Institute for Communications and Navigation, Department of Electrical and Computer Engineering,*
Technische Universität München, Munich, Germany
Email: {chen.zhu, gabriele.giorgi}@tum.de

[2] *Institute of Communications and Navigation,*
German Aerospace Center (DLR), Oberpfaffenhofen, Germany
Email: christoph.guenther@dlr.de

BIOGRAPHY

Mr. Chen Zhu is currently pursuing his Ph.D. at Technische Universität München as a full-time researcher at the Institute for Communications and Navigation. He received his B.Sc. in Automation Engineering from Tsinghua University, in Beijing, China in 2009, and his M.Sc. in Communications Engineering in 2011 from Technische Universität München, in Munich, Germany. His research interests include visual navigation, robotic swarm navigation, and multi-sensor fusion in autonomous vehicle navigation.

Dr. Gabriele Giorgi is a researcher at the Institute for Communications and Navigation, Technische Universität München, in Munich, Germany. He obtained a Ph.D. following his work on Global Navigation Satellite System (GNSS) for aerospace applications from the Delft Institute of Earth Observation and Space Systems (DEOS), Delft University of Technology, in Delft, The Netherlands. He holds a M.Sc. degree in space engineering from the University of Rome La Sapienza and a B.Sc. degree in aerospace engineering from the same university. His main research focuses on satellite navigation, visual navigation and multi-sensor fusion.

Prof. Christoph Günther studied theoretical physics at the Swiss Federal Institute of Technology in Zurich, Switzerland. He received his diploma in 1979 and completed his Ph.D. in 1984. He worked on research in cryptography, coding, communication, and information theory with Asea Brown Boveri, Ascom and Ericsson. Since 2003, he is the Director of the Institute of Communication and Navigation at the German Aerospace Center (DLR). The institute employs around 120 scientists. Since 2004, Günther is additionally holding the Chair of Communication and Navigation at Technische Universität München (TUM). The focus of his research work is on navigation. At TUM, he and his team are developing algorithms for achieving a high accuracy. At DLR, the focus is on achieving high levels of integrity.

ABSTRACT

Cooperative swarms of robots equipped with cameras are robust against failures, and can explore GNSS (Global Navigation Satellite System)-denied environments efficiently. VSLAM (Visual Simultaneous Localization and Mapping) techniques have been developed in recent years to estimate the trajectory of vehicles and to simultaneously reconstruct the map of the environment using visual clues. Due to constraints on payload size, weight, and costs, many VSLAM applications must be based on a single camera. The associated monocular estimation of the trajectory and map is ambiguous by a scale factor. This work shows that by exploiting sparse range measurements between a pair of dynamic rovers in planar motion, the correct scale factors of both cameras and the relative position, as well as the relative attitude between the rovers, can be estimated.

INTRODUCTION

Autonomous robotic platforms are utilized in the exploration of extreme environments, e.g., extraterrestrial exploration or catastrophe rescues. In order to increase the system robustness against hazards in the missions, e.g., strike during landing, and to improve the exploration efficiency, we propose to use a robotic swarm including multiple autonomous units such as multicopters and ground rovers [1] [2]. Autonomous navigation of the swarm elements often relies on several sensors such as mobile receivers, Inertial Measurement Units (IMUs), laser scanners and, most substantially, cameras [3]. Due to constraints on size, weight, accommodation and costs in swarm elements, monocular cameras are used instead of stereo rigs in most cases. VSLAM techniques using monocular cameras have been developed in recent years to estimate the trajectory of vehicles and to simultaneously reconstruct the map of the environment. Klein and Murray developed the Parallel Tracking and Mapping (PTAM) algorithm [4], which divides the tracking and mapping into separate threads to accelerate the computation. Engel, Schöps and Cremers proposed a large scale dense SLAM algorithm using monocular cameras [5], which minimizes the photometric error instead of the feature reprojection error for reducing the computational costs and improving the performance. Another state-of-the-art approach is ORB-SLAM from Mur-Artal, Montiel and Tardós [6]. The method utilizes ORB (Oriented FAST and Rotated BRIEF) features [7] and a novel keyframe-based graph structure, to provide a robust real-time monocular SLAM solution even in large scale scenarios and relatively low frame rate. However, all these algorithms estimate the motion only up to a global scale.

A number of approaches have been considered for resolving the global scale ambiguity. Most of them use IMUs, see for example Nützi et al. [8] and Abeywardena et al. [9]. However, the inherent drift of IMUs is prone to introducing estimation biases. Therefore, we developed a method for estimating the global scales of a pair of dynamic rovers in planar motion, using sparse range measurements on a single ranging link. In the case of a swarm of robots, these measurements could be performed between any pair of swarm elements [10]. Strictly, the algorithms developed in this paper do not depend on the method of ranging. It can be adapted without restrictions, from radio-frequency-based ranging to other sources of ranging measurements, e.g., radar or lidar. In addition, the relative position and attitude between the two rovers can be estimated. Fig. 1 shows a scenario of two dynamic rovers equipped with monocular cameras and a ranging link between them. Utilizing the cooperation between the pair of vehicles, the scale problem in VSLAM of both monocular cameras can be solved by the proposed method. Additionally, the scale estimation problem is coupled with the estimates of the initial relative position and attitude. As a consequence, the relative pose between the two rovers can be obtained simultaneously.

The paper continues with the system model and a brief introduction of motion estimation in monocular-camera-based VSLAM. Subsequently, a method of scale and relative pose estimation of two cooperative rovers using monocular cameras and sparse range measurements is proposed. Then the simulation results for various geometries and noise conditions are provided. The conclusions are drawn from the analysis of the results.

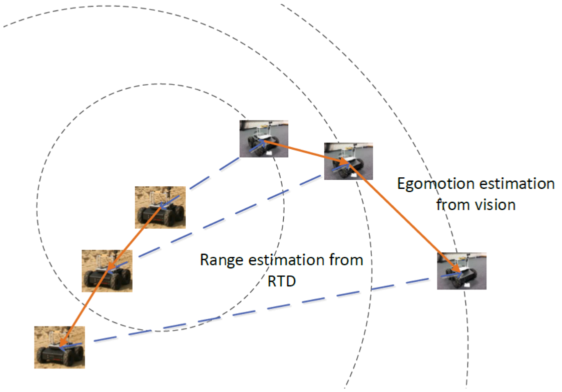


Figure 1: *Two dynamic rovers with ranging measurements*

SYSTEM MODEL AND MOTION ESTIMATION USING MONOCULAR CAMERAS

The measurement scenario addressed in this work is shown in Fig. 1. Two cooperative rovers equipped with a monocular camera and a ranging device, e.g., a wireless radio receiver, execute SLAM tasks on the ground. The motion of the vehicle is constrained to be planar. The additional difficulties of coping with the same problem in a three-dimensional (3D) motion are discussed in the following section. Let $\vec{c}_{[k]}^{(W)} \in \mathbb{R}^2$ be the position of the robot in world frame (W) at the time k. In the remainder of this paper, we use a superscript with parentheses $^{(\cdot)}$ to denote the coordinate frame in which the vector is represented. Vectors such as $\vec{c} \in \mathbb{R}^2$ with geometric meanings are written with an arrow. Time, denoted with square brackets $[\cdot]$, is measured in keyframes, i.e., the time reference instances in which both the range measurements and the trajectory estimation are available. We use (k) to express the local coordinate frame at keyframe k. The homogeneous coordinates in the extended Euclidean space are written as $\tilde{r} \in \mathbb{P}^2$. In addition, the origin of the body frame is defined at the position of the ranging sensor. Since the relative pose between the camera and the ranging sensor can be obtained by calibration, the body frame and camera frame are not distinguished. This assumption does not affect the validity of the algorithm if the body is assumed to be rigid.

The range measurements can be obtained by using pilot signals for synchronization. If the clock on the transmitter and receiver sides are precisely synchronized, the range can be estimated using time of arrival (ToA) measurements. If a satisfactory synchronization cannot be achieved, round-trip-delay (RTD) techniques can be implemented to eliminate the impact of the clock offset. The precision of the range measurements is constrained by their Cramér-Rao lower bound. The details of ranging using RTD for navigation purposes are discussed in [10].

In the proposed scheme, the rovers have basic communication capabilities so that one of them can transmit its local estimated trajectory $\{\vec{c}^{(N_1)}_{1,[k]}\}$ to the other one. The trajectory is estimated by VSLAM algorithm in the navigation frame of the rover, i.e., the fixed reference frame taking the starting location as the origin and the initial heading direction as the y-axis. Our method does not require transmission of extracted feature vectors or the local maps, so the data throughput requirement is significantly low. A radio-based system with both ranging and communication capabilities for robotic swarms is proposed by Zhang et. al. in [11].

To obtain the trajectory in navigation frame $\{\vec{c}^{(N)}_{[k]}\}$, the following steps of monocular-camera-based motion estimation are essential. Generally, the transformation between two coordinate frames (P) and (Q) follows

$$\vec{X}^{(Q)} = R_{(P\to Q)}\vec{X}^{(P)} + \vec{t}_{(P\to Q)}, \tag{1}$$

where $\vec{X}^{(P)}$ and $\vec{X}^{(Q)}$ denote the coordinates of an arbitrary 3D point $\vec{X} \in \mathbb{R}^3$ expressed in the corresponding (P) and (Q) frames, $R_{(P\to Q)} \in \mathbf{SO}(3)$ denotes the orthonormal rotation matrix, and $\vec{t}_{(P\to Q)}$ denotes the translation vector from the origin of (P) to the origin of (Q).

According to perspective projection, a visible point with 3D coordinates in the navigation frame $\vec{X}^{(N)}_i \in \mathbb{R}^3$ is projected to a two-dimensional (2D) point $\vec{u}^{(k)}_i$ in the measurement set $\Omega_{[k]}$ at k-th keyframe as

$$\vec{u}^{(k)}_i = \pi(\vec{X}^{(N)}_i, \vec{c}^{(N)}_{[k]}, R_{(N\to k)}) \in \Omega_{[k]} \subset \mathbb{R}^2. \tag{2}$$

$\Omega_{[k]}$ is the set consists of the 2D coordinates of all the points of interest on the image plane. In feature-based approaches, the measurement space Ω is continuous, whereas in direct methods it is a discrete set, i.e. the set of all the pixels. In homogenous coordinates, the projection can be simply denoted as

$$\tilde{u}^{(k)}_i = K P_{[k]} \tilde{X}^{(N)}_i, \tag{3}$$

where K denotes the camera intrinsic matrix, and $P_{[k]}$ the extrinsic projection matrix at time k. In planar motion case,

$$P_{[k]} = R_{(N\to k)} \left[\begin{array}{c|c} I_3 & \begin{matrix} -\vec{c}^{(N)}_{[k]} \\ 0 \end{matrix} \end{array}\right], \tag{4}$$

where I_3 denotes the three-dimensional identity matrix.

By tracking features in consecutive image sequences, the essential matrix $E_{(k\to k+1)}$ can be estimated using the epipolar geometry constraint:

$$(K^{-1}\tilde{u}^{(k+1)}_i)^T E_{(k\to k+1)} (K^{-1}\tilde{u}^{(k)}_i) = 0. \tag{5}$$

This equation shows the scale invariance of $E_{(k\to k+1)}$ in epipolar constraints. The essential matrix can be decomposed into a rotation $R_{(k\to k+1)}$ and a unit vector of translation $\vec{e}_{(k\to k+1)} \in \mathbb{R}^3$ as: $E_{(k\to k+1)} = \left[\vec{e}_{(k\to k+1)}\right]_\times R_{(k\to k+1)}$, where $[\cdot]_\times$ denotes the 3×3 skew symmetric matrix built as

$$\begin{bmatrix} e_1 \\ e_2 \\ e_3 \end{bmatrix}_\times = \begin{bmatrix} 0 & -e_3 & e_2 \\ e_3 & 0 & -e_1 \\ -e_2 & e_1 & 0 \end{bmatrix}. \tag{6}$$

The translation in true scale is related to the monocular estimation by

$$\vec{t}_{(k\to k+1)} = s_g l_{(k\to k+1)} \vec{e}_{(k\to k+1)}. \tag{7}$$

In this equation $l_{(k\to k+1)}\vec{e}_{(k\to k+1)}$ is the estimated translation from monocular vision, in which $l_{(k\to k+1)} \in \mathbb{R}^+$ denotes the estimated norm of the translation from time k to $k+1$, and $\vec{e}_{(k\to k+1)}$ denotes the direction of the motion. $s_g \in \mathbb{R}^+$ is the true global scale in the world frame, which cannot be obtained in the monocular-only case [12]. The relative scale between two translations can be estimated. Without loss of generality, one can assume $l_{(1\to 2)} = 1$. The 3D coordinates of the tracked points can be estimated by triangulation to build a local map. Consequently, a local optimization, e.g., bundle adjustment [13], shall be applied using the estimated motion to initialize the tracking thread of the SLAM algorithm. Then the positions at the following time instances can be obtained by minimizing the reprojection residual (photometric residual in direct method cases)

$$\hat{c}^{(N)}_{[k]} = \arg\min_{\vec{c}^{(N)}_{[k]}} \sum_{\vec{u}^{(k)}_i \in \Omega_{[k]}} \left\| \pi(\vec{X}^{(N)}_i, \vec{c}^{(N)}_{[k]}) - \vec{u}^{(k)}_i \right\|^2_{\Sigma^{-1}}, \tag{8}$$

where Σ is the measurements covariance matrix.

SCALE AND RELATIVE POSE ESTIMATION EXPLOITING SPARSE RANGE MEASUREMENTS

Without any other anchor point with known absolute position, one can only estimate the position and attitude of the cameras with respect to a known point in the navigation frame. We choose the initial position of the camera projection center of rover 2 as the coordinate reference system's origin, and the camera's principal axis as the y-axis. Fig. 2 illustrates the reference system and the geometry of the two rovers. The initial position and attitude of the two rovers can be expressed in the reference frame as

$$\vec{c}^{(W)}_{1,[1]} = r_1 R(\alpha)[1,0]^T,\ R_{(N_1\to W)} = R(\alpha + \theta - \frac{\pi}{2}). \tag{9}$$

$$\vec{c}^{(W)}_{2,[1]} = [0,0]^T, \quad R_{(N_2\to W)} = I_2, \tag{10}$$

where I_2 denotes the two-dimensional identity matrix, and $R(\cdot) \in \mathbf{SO}(2)$ denotes a 2D rotation matrix.

Using the images from the monocular cameras, the egomotion of the two rovers in their navigation frames can be independently estimated up-to-scale as $\{\vec{c}^{(N_1)}_{1,[k]}\}$ and $\{\vec{c}^{(N_2)}_{2,[k]}\}$.

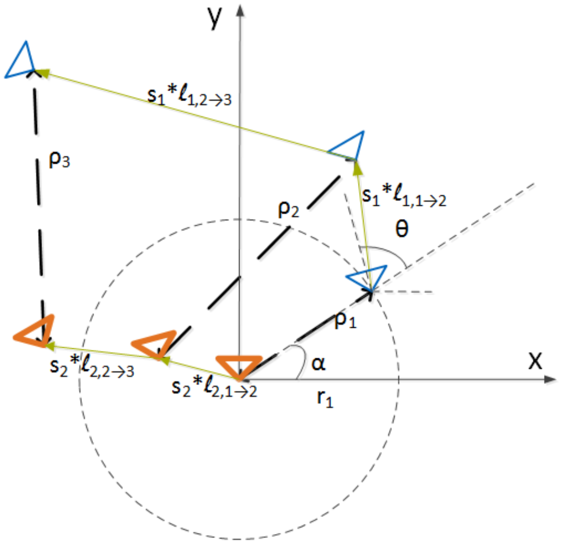


Figure 2: *Reference system and the geometry of the two rovers*

In the common reference frame (W), the position of the two rovers at k-th keyframe can be expressed as

$$\vec{c}^{(W)}_{1,[k]} = s_{g1} R_{(N_1 \to W)} \vec{c}^{(N_1)}_{1,[k]} + \vec{c}^{(W)}_{1,[1]} \tag{11}$$

$$\vec{c}^{(W)}_{2,[k]} = s_{g2} \vec{c}^{(N_2)}_{2,[k]} \tag{12}$$

Although the monocular camera itself can only estimate the motion with a scale ambiguity, with the additional help of a sparse set of noisy range measurements $\{\rho_k\}$, where

$$\rho_k = \left\| \vec{c}^{(W)}_{1,[k]} - \vec{c}^{(W)}_{2,[k]} \right\| + \eta_k, \tag{13}$$

a method for estimating the scale factors s_{g1}, s_{g2} can be devised by exploiting consecutive ranging measurements at keyframes. The true range between the two rovers at time k is

$$\begin{aligned} G_k(s_{g1}, s_{g2}, \alpha, \theta, r_1) &= \left\| \vec{c}^{(W)}_{1,[k]} - \vec{c}^{(W)}_{2,[k]} \right\| \\ &= \left\| s_{g1} R(\alpha + \theta - \frac{\pi}{2}) \vec{c}^{(N_1)}_{1,[k]} + r_1 R(\alpha)[1,0]^T - s_{g2} \vec{c}^{(N_2)}_{2,[k]} \right\|, \end{aligned} \tag{14}$$

which is determined by the rover trajectories in navigation frames and 5 unknown scalar parameters: the scale factors $s_{g1}, s_{g2} \in \mathbb{R}^+$,the polar angle $\alpha \in [0, 2\pi)$, the attitude angle $\theta \in [0, 2\pi)$, and the initial distance $r_1 \in \mathbb{R}^+$. We stack them into a parameter vector $\xi = [s_{g1}, s_{g2}, \alpha, \theta, r_1]^T$.

Utilizing communication functionality of the radio link between the two rovers, rover 1 can transmit its estimated motion (up-to-scale) to rover 2. Rover 2 serves as the master that obtains both local trajectory estimates. Therefore, by using the available set of range measurements, the unknown parameters can be estimated by minimizing

$$[\hat{\xi}] = \arg\min_{\xi} \|\rho - G(\xi)\|^2_{Q^{-1}}, \quad \text{s.t. } B\xi > 0, \tag{15}$$

where the vector $\rho = [\rho_1, \rho_2, ..., \rho_K]^T$ and $G(\xi) = [G_1(\xi), G_2(\xi), ..., G_K(\xi)]^T$.

$B = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$ is a selection matrix used to impose the positiveness of both scales and the initial distance. Q is the covariance matrix of the noise $\eta = [\eta_1, \eta_2, ..., \eta_K]^T$. The covariance of the range measurements can be obtained from the signal receiver. The Cramér-Rao lower bound of the range estimation can be used as an approximation when the covariance calculation is unavailable. If the ranging noise are uncorrelated, Q is a diagonal matrix.

Due to the bounded search space and the presence of several local minima, it is challenging to solve the nonlinear inequality constrained optimization in Eq. (15). However, not all minima violating the constraints represent erroneous solution, due to the symmetric properties of the objective function. According to Eq. (14), the norm G_k is invariant if the vector $\vec{c}^{(W)}_{1,[k]} - \vec{c}^{(W)}_{2,[k]}$ is reversed in direction. Consequently, for any parameter vector ξ, the value of the object function is invariant to the following parameter change:

$$\begin{aligned} & G_k(s_{g1}, s_{g2}, \alpha, \theta, r_1) \\ =& G_k(-s_{g1}, s_{g2}, \alpha, \theta + \pi, r_1) \\ =& G_k(-s_{g1}, -s_{g2}, \alpha + \pi, \theta + \pi, r_1) \\ =& G_k(-s_{g1}, -s_{g2}, \alpha, \theta, -r_1) \\ =& G_k(-s_{g1}, s_{g2}, \alpha + \pi, \theta, -r_1) \\ =& G_k(s_{g1}, -s_{g2}, \alpha + \pi, \theta, r_1) \\ =& G_k(s_{g1}, -s_{g2}, \alpha, \theta + \pi, -r_1) \\ =& G_k(s_{g1}, s_{g2}, \alpha + \pi, \theta + \pi, -r_1). \end{aligned} \tag{16}$$

Therefore, due to the numerical symmetry property of the cost function, we can obtain the estimates of the parameters by solving the corresponding unconstrained problem and transform the results obtained with the relations given in Table 1.

The nonlinear optimization problem (15) can be linearized to an unconstrained linearized least-squares problem

$$\hat{\xi} = \arg\min_{\xi} \|\rho - J(\xi)\xi\|^2_{Q^{-1}}, \tag{17}$$

with Jacobian matrix

$$J(\xi) = \begin{bmatrix} \frac{\partial G_1(\xi)}{\partial s_{g1}} & \frac{\partial G_1(\xi)}{\partial s_{g2}} & \frac{\partial G_1(\xi)}{\partial \alpha} & \frac{\partial G_1(\xi)}{\partial \theta} & \frac{\partial G_1(\xi)}{\partial r_1} \\ \frac{\partial G_2(\xi)}{\partial s_{g1}} & \frac{\partial G_2(\xi)}{\partial s_{g2}} & \frac{\partial G_2(\xi)}{\partial \alpha} & \frac{\partial G_2(\xi)}{\partial \theta} & \frac{\partial G_2(\xi)}{\partial r_1} \\ \vdots & & & & \\ \frac{\partial G_K(\xi)}{\partial s_{g1}} & \frac{\partial G_K(\xi)}{\partial s_{g2}} & \frac{\partial G_K(\xi)}{\partial \alpha} & \frac{\partial G_K(\xi)}{\partial \theta} & \frac{\partial G_K(\xi)}{\partial r_1} \end{bmatrix}.$$

Table 1: *Transformation on the results from unconstrained optimization.*

If			Transformation				
$\hat{s}_{g1} > 0$	$\hat{s}_{g2} < 0$	$\hat{r}_1 > 0$	$\hat{s}_{g1} \leftarrow \hat{s}_{g1}$	$\hat{s}_{g2} \leftarrow -\hat{s}_{g2}$	$\hat{\alpha} \leftarrow \hat{\alpha} + \pi$	$\hat{\theta} \leftarrow \hat{\theta}$	$\hat{r}_1 \leftarrow \hat{r}_1$
$\hat{s}_{g1} > 0$	$\hat{s}_{g2} < 0$	$\hat{r}_1 < 0$	$\hat{s}_{g1} \leftarrow \hat{s}_{g1}$	$\hat{s}_{g2} \leftarrow -\hat{s}_{g2}$	$\hat{\alpha} \leftarrow \hat{\alpha}$	$\hat{\theta} \leftarrow \hat{\theta} + \pi$	$\hat{r}_1 \leftarrow -\hat{r}_1$
$\hat{s}_{g1} > 0$	$\hat{s}_{g2} > 0$	$\hat{r}_1 < 0$	$\hat{s}_{g1} \leftarrow \hat{s}_{g1}$	$\hat{s}_{g2} \leftarrow \hat{s}_{g2}$	$\hat{\alpha} \leftarrow \hat{\alpha} + \pi$	$\hat{\theta} \leftarrow \hat{\theta} + \pi$	$\hat{r}_1 \leftarrow -\hat{r}_1$
$\hat{s}_{g1} < 0$	$\hat{s}_{g2} > 0$	$\hat{r}_1 > 0$	$\hat{s}_{g1} \leftarrow -\hat{s}_{g1}$	$\hat{s}_{g2} \leftarrow \hat{s}_{g2}$	$\hat{\alpha} \leftarrow \hat{\alpha}$	$\hat{\theta} \leftarrow \hat{\theta} + \pi$	$\hat{r}_1 \leftarrow \hat{r}_1$
$\hat{s}_{g1} < 0$	$\hat{s}_{g2} < 0$	$\hat{r}_1 > 0$	$\hat{s}_{g1} \leftarrow -\hat{s}_{g1}$	$\hat{s}_{g2} \leftarrow -\hat{s}_{g2}$	$\hat{\alpha} \leftarrow \hat{\alpha} + \pi$	$\hat{\theta} \leftarrow \hat{\theta} + \pi$	$\hat{r}_1 \leftarrow \hat{r}_1$
$\hat{s}_{g1} < 0$	$\hat{s}_{g2} < 0$	$\hat{r}_1 < 0$	$\hat{s}_{g1} \leftarrow -\hat{s}_{g1}$	$\hat{s}_{g2} \leftarrow -\hat{s}_{g2}$	$\hat{\alpha} \leftarrow \hat{\alpha}$	$\hat{\theta} \leftarrow \hat{\theta}$	$\hat{r}_1 \leftarrow -\hat{r}_1$
$\hat{s}_{g1} < 0$	$\hat{s}_{g2} > 0$	$\hat{r}_1 < 0$	$\hat{s}_{g1} \leftarrow -\hat{s}_{g1}$	$\hat{s}_{g2} \leftarrow \hat{s}_{g2}$	$\hat{\alpha} \leftarrow \hat{\alpha} + \pi$	$\hat{\theta} \leftarrow \hat{\theta}$	$\hat{r}_1 \leftarrow -\hat{r}_1$

The optimization (17) can be solved iteratively as

$$\hat{\xi}_{i+1} = \hat{\xi}_i + \left(J^T(\hat{\xi}_i)Q^{-1}J(\hat{\xi}_i)\right)^{-1} J^T(\hat{\xi}_i)Q^{-1}\left(\rho - G(\hat{\xi}_i)\right). \tag{18}$$

If s_{g1}, s_{g2} or r_1 are negative in the result, the parameters can be mapped to the symmetric solution in the valid search space as shown in Table 1. Consequently, the scales of the trajectories s_{g1} and s_{g2} are resolved. Additionally, the initial relative position and attitude between the two rovers are obtained. Combining with the trajectory estimates in navigation frames, the relative pose at any keyframe k can be estimated. As a distributed system, the master rover can transmit the estimation results to the other one using the communication system.

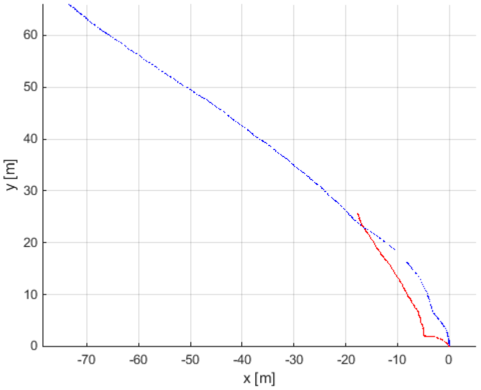


Figure 3: *Trajectories of the two rovers in Scenario #1.*

In order to solve the problem in Eq. (17), $K \geq 5$ range measurements are required. Due to the high nonlinearity of the objective function, the Levenberg-Marquardt algorithm is applied, instead of a Gauss-Newton approach, in order to exploit its superior global minimization capabilities. In addition, the initialization of the optimization is crucial due to the presence of a number of local minima. Although a suboptimal solution may have similar residual as the global minimum, the estimated parameters can be far away from the true value, leading to a wrong scale or pose. While ρ_1 is a precise approximation of the initial range r_1 due to the high accuracy of ranging measurements, the scaling factor s_{g1} and s_{g2} are significantly insensitive to the global minima problem, provided that the selected keyframes are sufficiently spaced. The estimation of the polar angle α and the attitude angle θ presents larger difficulties. Fortunately, the parameters to be estimated are constants and in most cases they do not need to be updated at high frequency. Hence a serial search for the proper initialization of the two angles is feasible. It is remarkable that if the relative position between the two rovers can be estimated by other methods, e.g., using ranging measurements from the swarm network in [10], the polar angle α could be precisely initialized. As a result, the search space would even reduce to a one-dimensional set. The requirement in initialization also explains why the problem with 3D motion is much more challenging. First of all, the local attitude angle have to be expressed with three orientation parameters and the initial relative position need to be parameterized by elevation and azimuth instead of α. Moreover, the objective function will have numerous local minima. As a result, with limited computational power, it is extremely challenging to obtain correct parameters in the 3D motion setup.

SIMULATION RESULTS

We test the proposed method on multiple trajectories using simulation data with Gaussian additive noise. The trajectories are generated with random walk processes as accelerations, starting from static locations with random relative position and attitude. In the simulation, there are two noise sources added on the measurements, the ranging noise with standard deviation σ_ρ and the translation noise with σ_t. In order to simulate a realistic scenario, the error of the trajectory estimation is added on all the translation estimates instead of on positions, i.e., the error accumulates over frames.

For the trajectories shown in Fig. 3, the root-mean-square-error (RMSE) of the parameter estimation under different noise levels is shown in Table 2. All the RMSE are calculated with ten repetitive runs with independent noise. Rover 2, i.e., the master node, is plotted in blue and rover 1 is in red. Fig. 4 shows the first 30 frames of the rovers to illustrate the initial relative geometry more clearly. In the serial search of initial values of the polar angle α and attitude angle θ, the grid size is set to be 10 degrees in the simulation. It can be concluded from the results that

Table 2: *Estimation error of scales and pose parameters in Scenario #1.*

σ_t	1 [cm]	1 [cm]	3 [cm]	5 [cm]	5 [cm]
σ_p	1 [cm]	10 [cm]	10 [cm]	10 [cm]	20 [cm]
$RMSE(s_{g1})$	0.0016	0.0049	0.0120	0.0129	0.0086
$RMSE(s_{g2})$	0.0015	0.0045	0.0127	0.0116	0.0067
$RMSE(\alpha)$ [deg]	3.3893	3.3426	8.2128	6.9004	8.8601
$RMSE(\theta)$ [deg]	0.6539	1.8069	4.2246	8.5225	6.2905
$RMSE(r_1)$ [m]	0.0171	0.0301	0.0596	0.1620	0.0716

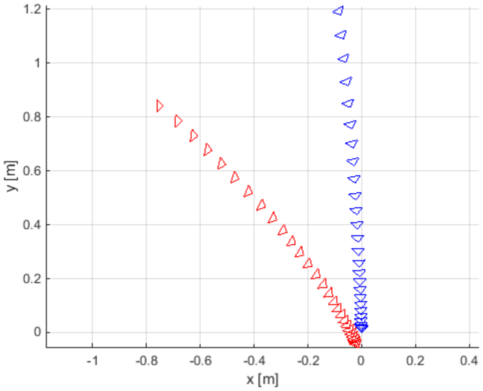


Figure 4: *First 30 frames of the two rovers in Scenario #1.*

in this scenario, the optimization converges well even for 5 [cm] translation error and 20 [cm] ranging noise. The scale factor of the trajectories for both rovers can be accurately estimated. An improvement in estimation precision of the angles can be obtained by setting a higher density of serial search values in the initialization of the non-linear optimization.

Other scenarios are also simulated to test the performance of the proposed method in different motion geometries. The trajectories of the rovers in various scenarios are shown in Fig. 5 and the corresponding estimation results are given in Table 3. It can be concluded that the method performs well in various scenarios with different geometries. A key factor that affects the precision of the estimation is the magnitude of the simulated motion. If the change of distance between the two rovers is comparable to the ranging noise, the measurement noise would be dominant in the estimation.

CONCLUSION

In many vision applications, a single camera is preferred over a stereo rig due to weight and cost constraints. However, the global scale is not recoverable in monocular vision. We propose an algorithm to resolve the global scale ambiguity in monocular VSLAM for a pair of cameras mounted on two rovers moving independently on a plane. By exploiting range measurements between the two rovers, the correct scales of the egomotions are estimated. At the same time, the relative position and attitude can be obtained. The algorithm was successfully tested on a number of simulated datasets with various geometries and noise patterns.

ACKNOWLEDGMENT

The project VaMEx-CoSMiC is supported by the Federal Ministry for Economic Affairs and Energy on the basis of a decision by the German Bundestag, grant 50NA1521 administered by DLR Space Administration.

REFERENCES

[1] Y. Mohan and S. G. Ponnambalam, “An extensive review of research in swarm robotics,” in *Nature Biologically Inspired Computing, 2009. NaBIC 2009. World Congress on*, Dec. 2009, pp. 140–145.

[2] S. Sand, S. Zhang, M. Mühlegg, G. Falconi, C. Zhu, T. Krüger, and S. Nowak, “Swarm exploration and navigation on Mars,” in *International Conference on Localization and GNSS, Torino, Italy*, 2013.

[3] M. Maimone, Y. Cheng, and L. Matthies, “Two years of visual odometry on the mars exploration rovers,” *Journal of Field Robotics*, vol. 24, no. 3, pp. 169–186, 2007.

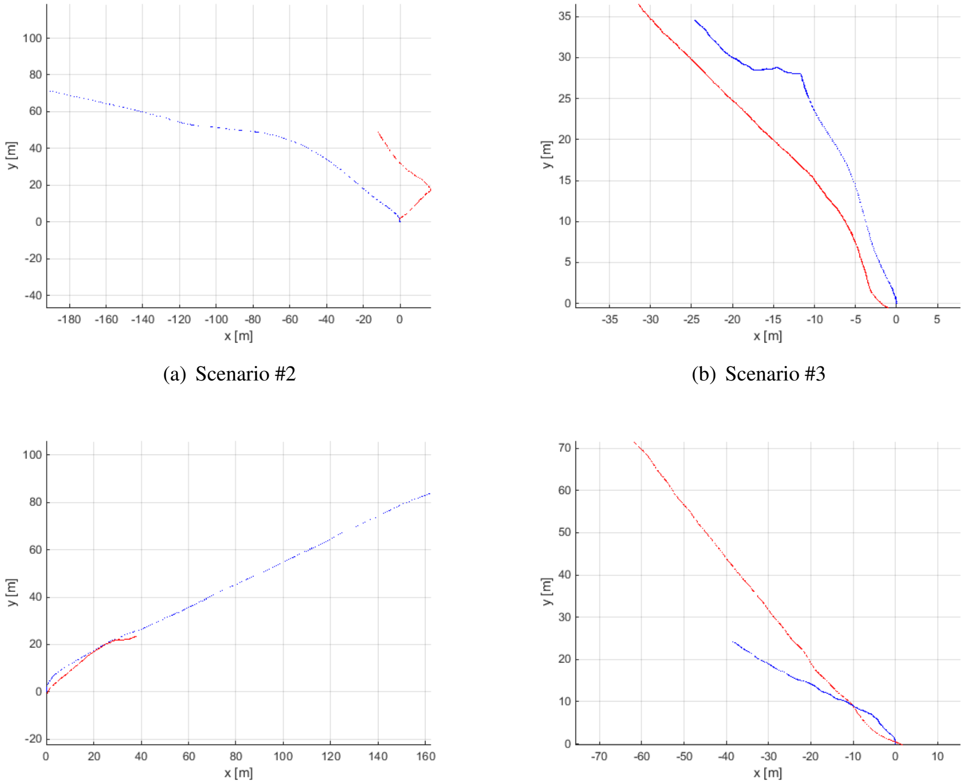
[4] G. Klein and D. Murray, “Parallel tracking and mapping for small ar workspaces,” in *Mixed and Augmented Reality, 2007. ISMAR 2007. 6th IEEE and ACM International Symposium on*, Nov 2007, pp. 225–234.

[5] J. Engel, T. Schöps, and D. Cremers, “Lsd-slam: Large-scale direct monocular slam,” in *Computer Vision - ECCV 2014*, ser. Lecture Notes in Computer Science, D. Fleet, T. Pajdla, B. Schiele, and T. Tuytelaars, Eds. Springer International Publishing, 2014, vol. 8690, pp. 834–849.

[6] R. Mur-Artal, J. M. M. Montiel, and J. D. Tardos, “Orb-slam: A versatile and accurate monocular slam system,” *IEEE Transactions on Robotics*, vol. 31, no. 5, pp. 1147–1163, Oct 2015.

Table 3: *Estimation error of parameters in various scenarios.*

Scenario #	2	3	4	5
σ_t	5 [cm]	5 [cm]	3 [cm]	1 [cm]
σ_p	20 [cm]	10 [cm]	10 [cm]	5 [cm]
$RMSE(s_{g1})$	0.0052	0.0030	0.0149	0.0009
$RMSE(s_{g2})$	0.0018	0.0017	0.0048	0.0033
$RMSE(\alpha)$ [deg]	2.9416	9.7634	9.0721	5.8405
$RMSE(\theta)$ [deg]	3.5057	6.0490	8.6343	0.7336
$RMSE(r_1)$ [m]	0.1337	0.0717	0.0486	0.0205



(a) Scenario #2

(b) Scenario #3

(c) Scenario #4

(d) Scenario #5

Figure 5: *Trajectories of the two rovers in various scenarios.*

[7] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski, “Orb: an efficient alternative to sift or surf,” in *Computer Vision (ICCV), 2011 IEEE International Conference on*. IEEE, 2011, pp. 2564–2571.

[8] G. Nützi, S. Weiss, D. Scaramuzza, and R. Siegwart, “Fusion of imu and vision for absolute scale estimation in monocular slam,” *Journal of Intelligent and Robotic Systems*, vol. 61, no. 1-4, pp. 287–299, 2011.

[9] D. Abeywardena, Z. Wang, S. Kodagoda, and G. Dissanayake, “Visual-inertial fusion for quadrotor micro air vehicles with improved scale observability,” in *Robotics and Automation (ICRA), 2013 IEEE International Conference on*, May 2013, pp. 3148–3153.

[10] E. Staudinger, S. Zhang, A. Dammann, and C. Zhu, “Towards a radio-based swarm navigation system on mars - key technologies and performance assessment,” in *Wireless for Space and Extreme Environments (WiSEE), 2014 IEEE International Conference on*, Oct 2014, pp. 1–7.

[11] S. Zhang, S. Sand, R. Raulefs, and E. Staudinger, “Self-organized hybrid channel access method for an interleaved RTD-based swarm navigation system,” in *Workshop on Positioning, Navigation and Communication, Dresden, Germany*, 2013.

[12] D. Scaramuzza and F. Fraundorfer, “Visual odometry [tutorial],” *Robotics & Automation Magazine, IEEE*, vol. 18, no. 4, pp. 80–92, 2011.

[13] B. Triggs, P. F. McLauchlan, R. I. Hartley, and A. W. Fitzgibbon, “Bundle adjustmenta modern synthesis,” in *Vision algorithms: theory and practice*. Springer, 1999, pp. 298–372.