Planar Pose Estimation using a Camera and Single-Station Ranging 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

Cameras are widely used for localization and navigation in GNSS-denied environments. By exploiting VSLAM (Visual Simultaneous Localization and Mapping) techniques, vehicles equipped with cameras are capable of estimating their own trajectories and simultaneously building a map of the surrounding environment. Due to constraints on payload size, weight, and costs, many VSLAM applications must be based on a single camera. However, the associated monocular estimation of the vehicle trajectory and the map is ambiguous by a scale factor. The purpose of this work is to show how the correct scale factor can be estimated in planar motion cases by exploiting range measurements from a single station. The proposed method is independent of the VSLAM algorithm used for ego-motion estimation of the vehicle.

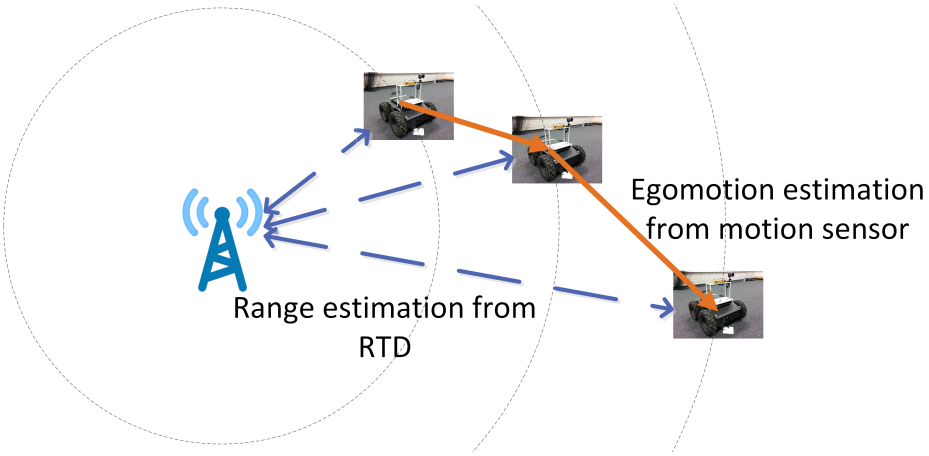

Figure 1: Static station and dynamic rover

INTRODUCTION

Autonomous navigation of ground vehicles often relies on several sensors such as mobile receivers, Inertial Measurement Units (IMUs), laser scanners and cameras [1]. By applying VSLAM (Visual Simultaneous Localization and Mapping) techniques, a vehicle can estimate its ego-motion and simultaneously build a map using onboard cameras. Due to constraints on size, weight, accommodation and cost, a stereo camera rig cannot be implemented in many cases. Moreover, if the baseline length of a stereo rig is very short, it does not provide significant advantage over a single camera due to the resolution limits. As a result, several approaches using monocular cameras have been developed. Klein and Murray developed the Parallel Tracking and Mapping (PTAM) algorithm [2], which divides the tracking and mapping into separate threads to accelerate the computation. Later on, Strasdat, Montiel and Davison proposed a scale-aware loop closure method to deal with the relative scene scale drift [3]. Based on the PTAM framework, Mur-Artal, Montiel and Tardós [4] proposed the state-of-the-art ORB-SLAM approach, which provides a robust real-time monocular SLAM solution. Engel, Schöps and Cremers proposed a large scale dense SLAM algorithm (LSD-SLAM) using monocular cameras [5], which minimizes the photometric error instead of the feature reprojection error for improving the performance. 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. Many of them use IMUs, see for example Achtelik et al. [6] and Nützi et al. [7]. The fusion of camera and IMUs can significantly reduce the relative scale drifts. However, the inherent drift of IMUs is prone to introducing inconsistency in global scale estimation. Tabibiazar and Basir proposed an approach using range measurements from cellular networks [8]. Their method requires at least 3 ranging links for estimating the robot's position before integrating the result with vision-based estimates. Three links are often not available, e.g., mobile networks do not typically provide threefold coverage. Zhu, Giorgi, and Günther proposed a scale and relative position estimation method in [9] using two dynamic rovers equipped with monocular cameras and ranging capability between them. It provides a solution for multi-agent based applications, e.g., robotic swarm navigation. Nevertheless, the cooperation of two dynamic vehicles is not always feasible.

Therefore, we developed a method for estimating the global scale in monocular VSLAM by exploiting range measurements on a single link to a static station. As a result, the The method developed in this paper does not depend on the method of ranging, as long as it is performed with respect to a given static location.

SYSTEM MODEL

The measurement scenario addressed in this work is shown in Fig. 1. A rover equipped with a monocular camera and a ranging device, e.g., a wireless radio receiver, executes SLAM tasks on the ground. The motion of the vehicle is constrained to be planar. In the proposed scheme, a radio link is available between the dynamic rover and a static station. The static station can be another robotic rover in stationary mode, a base station, or a wireless hotspot with transmitter for the specific radio signal. The dynamic rover estimate its ego-motion using an onboard camera, and a radio link is established to execute base-to-rover ranging. The range measurements can be obtained by using pilot signals for synchronization. Because a satisfactory clock synchronization between the rover and the station cannot be achieved in many cases, round-trip-delay (RTD) techniques is a favorable choice for slow-movement scenarios to eliminate the impact of any clock offset. The details of ranging using RTD for navigation purposes are discussed in [10].

We define a navigation frame (N) as a fixed coordinate frame for the rover with its origin at the starting location of rover. The navigation frame is related to the world reference frame by a specific transformation dependent on the initial position and

attitude of the vehicle. Moreover, we use (k) to express the camera's local coordinate frame at keyframe k, which varies as the camera moves. Let $\vec{c}_k^{(W)} \in \mathbb{R}^2$ be the position of the robot in world frame (W) at 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. Time is measured at keyframes, i.e., the time reference instances in which both the range measurements and the trajectory estimation are available.

The transformation between two arbitrary 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 $\vec{X}$ 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).

MOTION ESTIMATION USING MONOCULAR CAMERAS

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

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

Ω_k is the set of 2D coordinates of all the points of interest on the image plane.

By tracking features in consecutive image sequences, the essential matrix $E_{(k\to k+1)}$ can be estimated using epipolar geometry constraints [11]. The essential matrix can be decomposed into a rotation $R_{(k\to k+1)}$ and a unit vector $\vec{e}_{(k\to k+1)}$ representing the translation direction 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

$$\begin{bmatrix} e_1 \\ e_2 \\ e_3 \end{bmatrix}_\times = \begin{bmatrix} 0 & -e_3 & e_2 \\ e_3 & 0 & -e_1 \\ -e_2 & e_1 & 0 \end{bmatrix}. \tag{3}$$

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

$$\vec{t}_{(k\to k+1)} = s_g l_{(k\to k+1)}\vec{e}_{(k\to k+1)}. \tag{4}$$

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 $\vec{e}_{(k\to k+1)}$ reflects 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]. Without loss of generality, one can assume $l_{(1\to 2)} = 1$, since one can map the value into s_g.

To start the tracking, using the estimated motion from the first two frames, the 3D coordinates of the tracked points can be estimated by triangulation as $\hat{X}_i^{(N)} = \pi^{-1}(\vec{u}_i^{(1)}, \vec{u}_i^{(2)}, E_{(1\to 2)})$. Consequently, the camera positions at the following time instances $k \geq 2$ can be obtained by minimizing the re-projection residual

$$\hat{c}_{k+1}^{(N)} = \arg\min_{\vec{c}_{k+1}^{(N)}} \sum_{\vec{u}_i^{(k+1)} \in \Omega_{k+1}} \left\| \pi(\vec{X}_i^{(N)}, \vec{c}_{k+1}^{(N)}) - \vec{u}_i^{(k+1)} \right\|^2_{\Sigma^{-1}_{i,k+1}}, \tag{5}$$

where where $\|\cdot\|_{\Sigma^{-1}}$ denotes the Mahalanobis distance with Σ as the measurements covariance matrix. Using the estimated pose, the 3D position of the new features detected in frame $k+1$ can be updated using $\pi^{-1}(\cdot)$. As a result, the tracking can be continued as long as sufficient features can be tracked in consecutive frames. However, the estimated position of the camera and the map points are all ambiguous by a scale factor s_g.

During the consistent tracking, keyframes are selected if the image has significant change while sufficient reliable features are tracked. Since the obtained motion estimates follow dead-reckoning principle, the estimation error will accumulate over time. In order to improve the accuracy of the estimation result, a global optimization for both 3D point position and the vehicle poses is performed using K keyframes and N_p map points:

$$\{\hat{X}_i^{(N)}\}, \{\hat{c}_k^{(N)}\}, \{\hat{R}_{(N\to k)}\} = \arg\min_{\{\vec{X}_i^{(N)}, \vec{c}_k^{(N)}, R_{(N\to k)}\}} \sum_{i=1}^{N_p}\sum_{k=1}^{K} v_{ik} \left\| \vec{u}_i^{(k)} - \pi(\vec{X}_i^{(N)}, \vec{c}_k^{(N)}, R_{(N\to k)}) \right\|^2_{\Sigma^{-1}_{i,k}}, \tag{6}$$

where v_{ik} is a binary visibility mask, which assumes $v_{ik} = 1$ if feature i is visible to the camera at time instant k, otherwise $v_{ik} = 0$.

Therefore, by executing the optimization in Eqn. (6), the rover obtains a set of up-to-scale egomotion estimates expressed in its own navigation frame, i.e., $\{\hat{c}_k^{(N)}\}$.

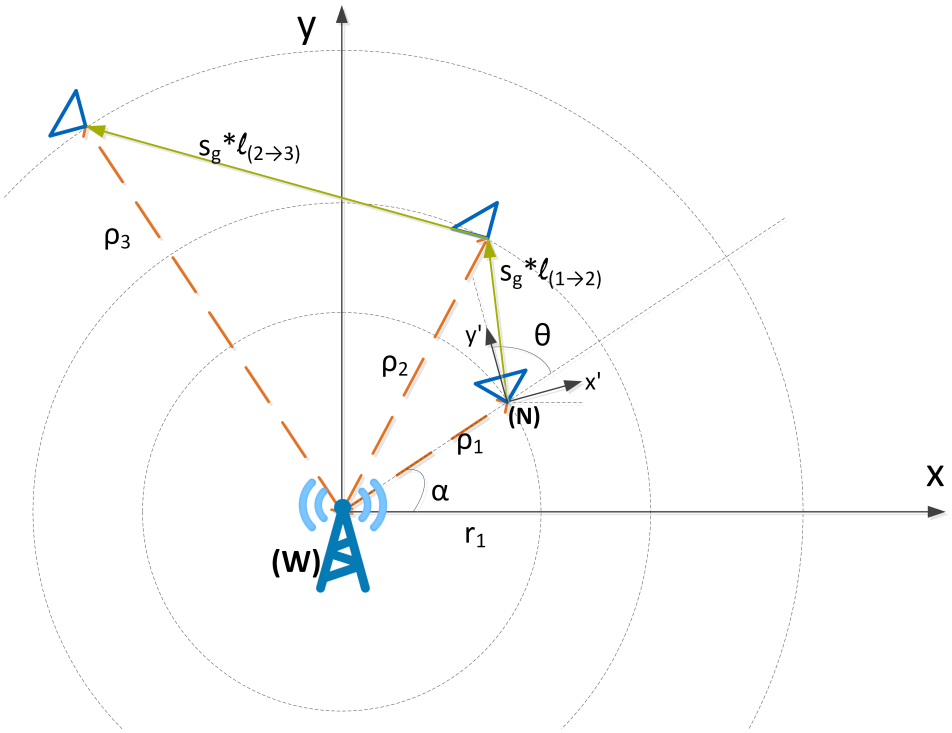

Figure 2: Geometry of the static station and the dynamic rover

SCALE ESTIMATION USING MONOCULAR CAMERA AND SINGLE RANGING LINK

Fig. 2 illustrates the basic geometry between the static station and the dynamic rover. Without loss of generality, the reference frame origin is set to be the position of the base station. The direction of the x-axis of the reference frame (W) can be arbitrarily chosen. The initial heading of the camera is defined as the y'-axis in the navigation frame (N). As a result, the initial position of the rover can be parameterized by the initial radius r_1 and the initial polar angle α in the world frame (W). The initial attitude of the rover is described by the angle $\alpha+\theta-\frac{\pi}{2}$, i.e.,

$$\vec{c}_1^{(W)} = t_{(N\to W)} = r_1 R(\alpha)[1,0]^T, \tag{7}$$

$$R_{(1\to W)} = R_{(N\to W)} = R(\alpha+\theta-\frac{\pi}{2}), \tag{8}$$

where $R(\cdot)\in\mathbf{SO}(2)$ is a 2D rotation matrix.

Using the images from the monocular cameras, the ego-motion of the rover in its navigation frame (N) can be independently estimated up-to-scale as $\{\hat{c}_k^{(N)}\}$ by exploiting a VSLAM algorithm. However, due to the lack of global scale knowledge, the poses in real-world metric are unavailable.

In the world reference frame (W), the position of the rover at k-th keyframe can be expressed as

$$\vec{c}_k^{(W)} = s_g R_{(N\to W)} \vec{c}_k^{(N)} + \vec{c}_1^{(W)} \tag{9}$$

As a result, the coordinates of the trajectory in world frame and navigation frame can be related by a similarity transformation $T\in\mathfrak{Sim}(2)\subset\mathbb{R}^{3\times3}$. The transformation in Eq. (9) can be parameterized by 4 parameters as:

$$\vec{c}_k^{(W)} = R(\alpha)\left(r_1[1,0]^T + s_g R(\theta-\frac{\pi}{2})\vec{c}_k^{(N)}\right) \tag{10}$$

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

$$\rho_k = r_k + \eta_k = \left\|\vec{c}_k^{(W)}\right\| + \eta_k, \tag{11}$$

a method for estimating the global scaling factor s_g can be devised by exploiting consecutive ranging measurements at keyframes.

Table 1: Transformation on the results from unconstrained optimization.

If		Transformation	
$\hat{s}_g < 0$	$\hat{r}_1 > 0$	$\hat{s}_g \leftarrow -\hat{s}_g$	$\hat{\theta} \leftarrow \hat{\theta} + \pi$
$\hat{s}_g > 0$	$\hat{r}_1 < 0$	$\hat{r}_1 \leftarrow -\hat{r}_1$	$\hat{\theta} \leftarrow \hat{\theta} + \pi$
$\hat{s}_g < 0$	$\hat{r}_1 < 0$	$\hat{s}_g \leftarrow -\hat{s}_g$	$\hat{r}_1 \leftarrow -\hat{r}_1$

Using the range measurements, the global scale s_g, initial attitude heading θ, and initial radius r_1 can be estimated by least-squares optimization. Stacking the K range measurements and the three parameters into vectors $\rho = [\rho_1, \rho_2, ..., \rho_K]^T$ and $F(\xi) = [\|\vec{c}_1^{(W)}\|, \|\vec{c}_2^{(W)}\|, ..., \|\vec{c}_K^{(W)}\|]^T$ with $\xi = [s_g, \theta, r_1]^T$, the problem can be formulated as

$$\hat{\xi} = \arg\min_{\xi} \|\rho - F(\xi)\|^2_{Q^{-1}} \quad \text{s.t. } B\xi > 0. \tag{12}$$

The inequality constraints are due to the positiveness of both the scale s_g and the initial true range r_1. $B = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ is a selection matrix used to set the constraints. Q is the covariance matrix of the noise $\eta = [\eta_1, \eta_2, ..., \eta_K]^T$.

Due to the presence of several local minima and the bounded search space, it is challenging to solve the nonlinear inequality constrained optimization in Eq. (12). However, not all minima violating the constraints represent erroneous solution, due to the symmetric properties of the objective function. Define $A_k(s_g, \theta, r_1) = (r_k(\xi) - \rho_k)^2$. For any s_g, θ and r_1, the value of object function is invariant to the following parameter change:

$$\begin{aligned} &A_k(s_g, \theta, r_1) = A_k(-s_g, \theta + \pi, r_1) \\ =&A_k(s_g, \theta + \pi, -r_1) = A_k(-s_g, \theta, -r_1). \end{aligned} \tag{13}$$

Consequently, we can obtain the estimates of the parameters by solving the corresponding unconstraint problem and transform the results obtained with the relations given in Table 1.

The unconstrained optimization can be obtained iteratively by solving a linearized problem as

$$\hat{\xi} = \arg\min_{\xi} \left\|\rho - J_\xi \xi\right\|^2_{Q^{-1}} \tag{14}$$

$$\hat{\xi}_{i+1} = \hat{\xi}_i + \left(J_\xi(\hat{\xi}_i)^T Q^{-1} J_\xi(\hat{\xi}_i)\right)^{-1} J_\xi(\hat{\xi}_i)^T Q^{-1} \left(\rho - F(\hat{\xi}_i)\right) \tag{15}$$

where J_ξ is the Jacobian matrix associated to the function $F(\xi)$.

In order to solve the problem in Eq. (14), $K \geq 3$ range measurements are required. Due to the high nonlinearity of the objective function, the Levenberg-Marquardt algorithm [13] is applied, instead of a Gauss-Newton approach [14], in order to exploit its better global minimization capabilities.

As a result, the global scale factor s_g is estimated by combining ranging and visual measurements. Consequently, based on the global scale, the position of the rover $\{\hat{c}_k^{(W)}\}$ and the coordinates of the map points $\{\vec{X}_i^{(W)}\}$ can be obtained without scale ambiguity. In addition, the initial heading of the rover θ, which refers to the radial direction between the rover and the station, can be obtained from the estimation. It should be mentioned that the ranges are invariant to the change of the initial polar angle α. Hence with a single radio link, the absolute position of the rover in reference frame (W) is ambiguous by the angle. The ambiguity can be resolved only if additional set-ups are available, e.g., by connecting to a second base station, or by using an antenna array to estimate the angle of arrival.

SIMULATION RESULTS

The proposed scale estimation method using sparse range measurements is tested in simulation using KITTI benchmark datasets [15]. We use the odometry dataset with provided ground truth to verify the result. The range measurements $\{\rho_k\}$ are simulated from the true ranges with additive Gaussian noise.

Fig. 3 shows the estimation result at keyframe 100 from KITTI odometry dataset 07. The image at the current keyframe is shown in the upper-left plot and the estimated trajectory in the navigation frame until current time instant is at upper-right. The uncertainty of the range measurements is 1 [m]. The figure at bottom-left compares the whole trajectory in estimated global scale with the ground truth data. The true trajectory is plotted in blue color, and the estimated one is in red. It can be observed that the scaled trajectory using the estimated $\hat{s}_g$ aligns well with the ground truth. The bottom-right curve reflects the change the scale estimation error over time. The result converges quickly and the error remains low after 20 keyframes. As a comparison, Fig.

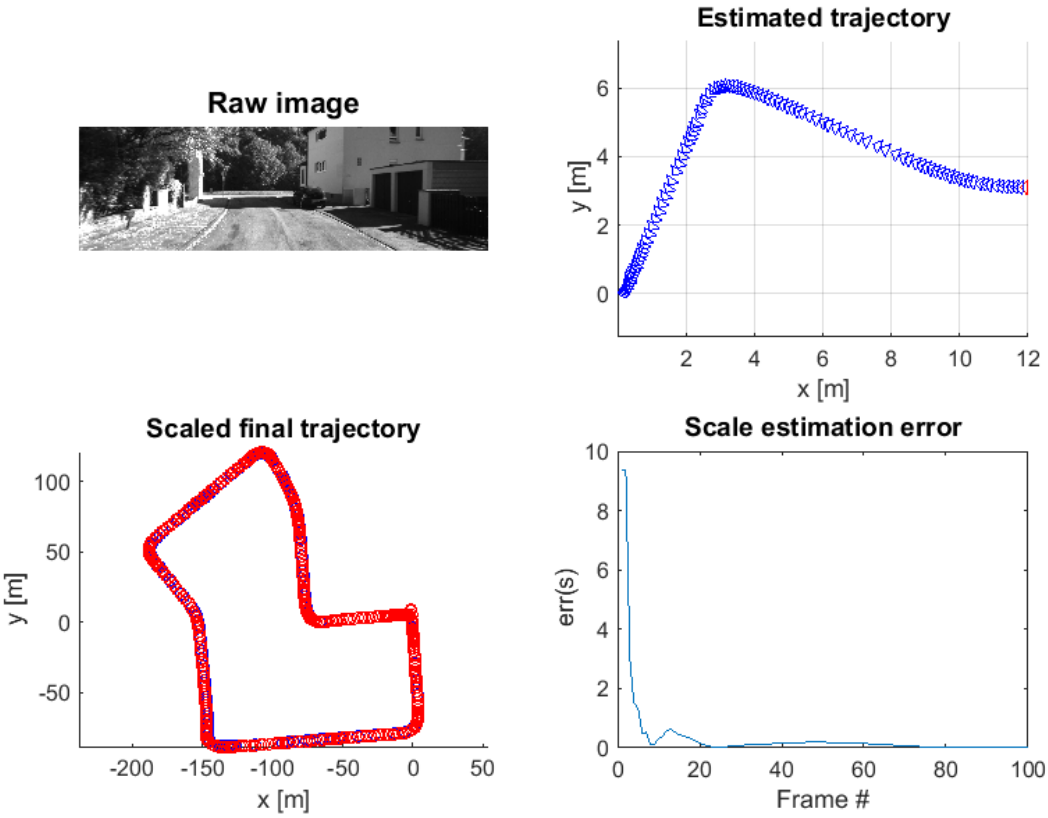

Figure 3: Scale estimation using KITTI odometry dataset 07

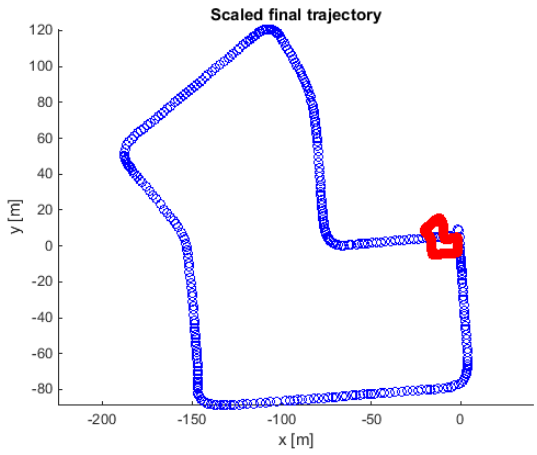

Figure 4: True scale and the initial guess for KITTI odometry dataset 07

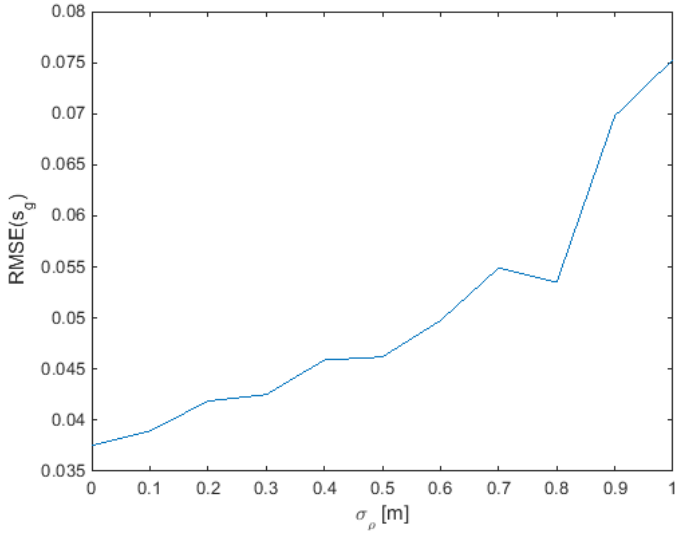

Figure 5: Impact of ranging error on the scale estimation error

4 illustrates the difference between the true global scale and the initial guess without any scale estimation. Since a monocular camera cannot provide any global scale information, the red trajectory is scaled using $s_g = 1$.

To analyze the impact of the ranging noise on the estimation result, we execute the simulation by adding different ranging noise on the measurements. Fig. 5 visualizes the result of the global scale estimation error with respect to the ranging noise level. The images used in the simulation are taken from KITTI dataset 07. The curve shows the root-mean-square error (RMSE) of global scale estimation with respect to the standard deviation of the range measurements. It can be seen that the scale estimation error is superlinearly dependent on the range measurements error. The curve is generated using KITTI dataset 07 with ground truth scale $s_g = 10.3624$. The relative error is less than 0.8% when the ranging uncertainty is below 1 meter. In practice, dependent on the requirements and constraints for the SLAM scenario, higher ranging accuracy than 1 meter is feasible through signal design.

Fig. 6, Fig. 7 and Fig. 8 shows the estimated poses of the rover using images from KITTI odometry dataset 04, 03 and 07 by applying the jointly estimated parameters $\hat{\xi}$. The dataset 04 contains images taken from a linear trajectory without any direction change. The three datasets represent different typical motions of a dynamic rover. The dataset 03 is a forward trajectory with a few turns, and the dataset 07 consists of images taken from motions in a closed loop. In the figures, the red trajectories are the ground truth, and the blue ones are the estimated trajectories. In the simulation, the standard deviation of the ranging noise is 1 meter. It can be conclude from the results that using the proposed method, the poses of the rover can be well estimated with only a single camera and sparse ranging measurements from a single base station.

CONCLUSION

For VSLAM applications based on monocular cameras, the estimated positions of the camera and the map points have a global scale ambiguity. We propose a method that exploits ranging measurements from a single station to estimate the scale factor in planar motion cases. Applying the symmetry property of the cost function, one can solve an unconstrained optimization while preserves the positiveness of the scale by using a transformation. The proposed method is verified on real images from benchmark datasets, and it is shown that the camera poses can be accurately estimated without global scale ambiguity.

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] 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.

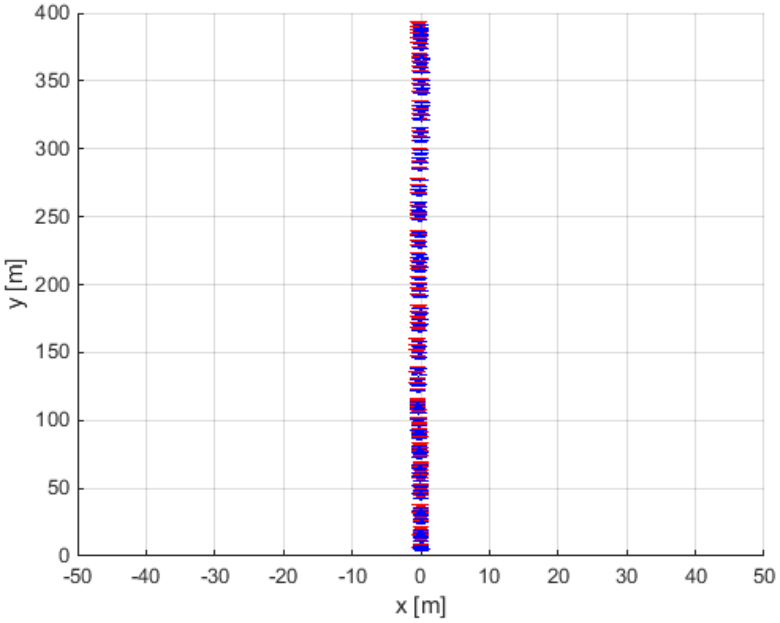

Figure 6: Pose estimation using KITTI odometry dataset 04

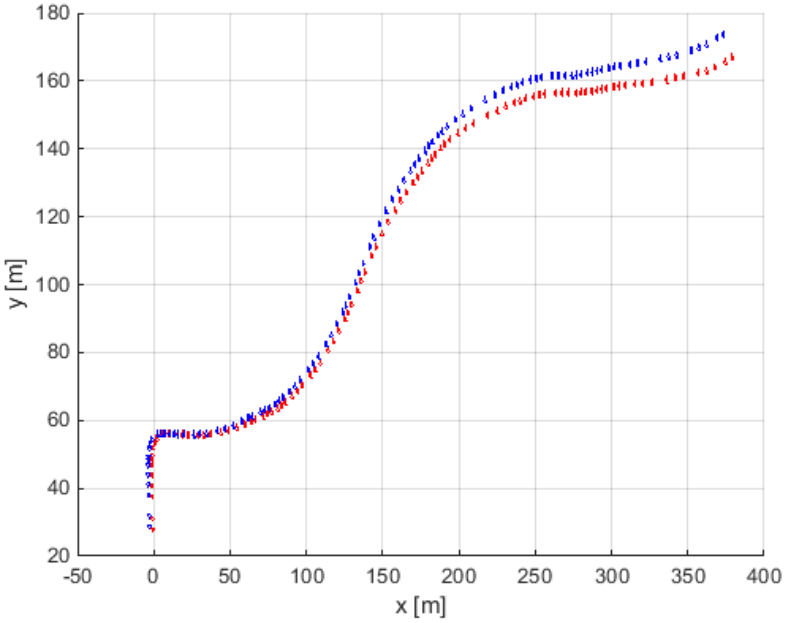

Figure 7: Pose estimation using KITTI odometry dataset 03

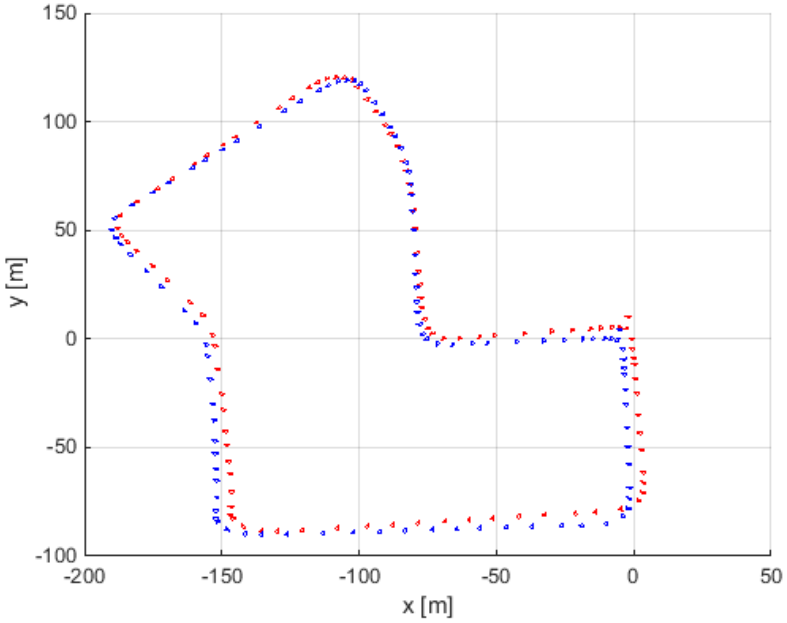

Figure 8: Pose estimation using KITTI odometry dataset 07

[2] 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.

[3] H. Strasdat, J. Montiel, and A. J. Davison, "Scale drift-aware large scale monocular slam." in *Robotics: Science and Systems*, vol. 2, no. 3, 2010, p. 5.

[4] 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.

[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] M. Achtelik, M. Achtelik, S. Weiss, and R. Siegwart, "Onboard imu and monocular vision based control for mavs in unknown in- and outdoor environments," in *Robotics and Automation (ICRA), 2011 IEEE International Conference on*, May 2011, pp. 3056–3063.

[7] 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.

[8] A. Tabibiazar and O. Basir, "Radio-visual signal fusion for localization in cellular networks," in *Multisensor Fusion and Integration for Intelligent Systems (MFI), 2010 IEEE Conference on*, Sept 2010, pp. 150–155.

[9] C. Zhu, G. Giorgi, and C. Günther, "Scale and 2d relative pose estimation of two rovers using monocular cameras and range measurements," in *Proceedings of the 29th International Technical Meeting of The Satellite Division of the Institute of Navigation (ION GNSS+ 2016), Portland, Oregon*. Institute of Navigation, 2016, pp. 794–800.

[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] R. Hartley and A. Zisserman, *Multiple view geometry in computer vision*. Cambridge university press, 2003.

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

[13] J. J. Moré, "The levenberg-marquardt algorithm: implementation and theory," in *Numerical analysis*. Springer, 1978, pp. 105–116.

[14] Y. Wang, "Gauss–newton method," *Wiley Interdisciplinary Reviews: Computational Statistics*, vol. 4, no. 4, pp. 415–420, 2012.

[15] A. Geiger, P. Lenz, and R. Urtasun, "Are we ready for autonomous driving? the kitti vision benchmark suite," in *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2012.