

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/328040203>

Fusion of Monocular Vision and Radio-based Ranging for Global Scale Estimation and Drift Mitigation

Preprint · October 2018

CITATIONS

0

READS

66

4 authors:



Young-Hee Lee

Technische Universität München

6 PUBLICATIONS 6 CITATIONS

[SEE PROFILE](#)



Chen Zhu

Technische Universität München

23 PUBLICATIONS 102 CITATIONS

[SEE PROFILE](#)



Gabriele Giorgi

German Aerospace Center (DLR)

62 PUBLICATIONS 679 CITATIONS

[SEE PROFILE](#)



Christoph Günther

Technische Universität München

56 PUBLICATIONS 382 CITATIONS

[SEE PROFILE](#)

Some of the authors of this publication are also working on these related projects:



Testing General Relativity Using Galileo Satellite Signals [View project](#)

Fusion of Monocular Vision and Radio-based Ranging for Global Scale Estimation and Drift Mitigation

Young-Hee Lee¹, Chen Zhu², Gabriele Giorgi² and Christoph Günther³

Abstract—Monocular vision-based Simultaneous Localization and Mapping (SLAM) is used for various purposes due to its advantages in cost, simple setup, as well as availability in the environments where navigation with satellites is not effective. However, camera motion and map points can be estimated only up to a global scale factor with monocular vision. Moreover, estimation error accumulates over time without bound, if the camera cannot detect the previously observed map points for closing a loop. We propose an innovative approach to estimate a global scale factor and reduce drifts in monocular vision-based localization with an additional single ranging link. Our method can be easily integrated with the back-end of monocular visual SLAM methods. We demonstrate our algorithm with real datasets collected on a rover, and show the evaluation results.

I. INTRODUCTION

Simultaneous Localization and Mapping (SLAM) with a single monocular camera is implemented for various applications for estimating camera motion and map points with a simple cost- and weight-effective setup. However, only relative scales between estimates can be recovered with monocular vision-based approaches, since no real scale information is available. Even though a calibrated stereo rig with a known baseline length can be a simple solution to resolve the global scale ambiguity, it is not an efficient tool for mapping distant environments, since the uncertainty in the depth estimation increases quadratically with the distances of scenes. Furthermore, the estimation error in monocular visual odometry accumulates over time without bound. The loop closing technique [1] is an efficient tool to correct the drifts. However, a camera has to observe the same scenes to close a loop, which becomes a significant constraint in mission design. In addition, it requires additional computational loads and data storage. Moreover, it is hard to verify the quality of loop detection.

We propose a fusion of monocular visual odometry and sparse radio-based ranging to resolve global scale ambiguity and reduce drifts. Our approach exploits a single ranging-anchor, hence the number of ranging measurements increases

only linearly with time. In addition, a radio signal can cover large areas ($\sim 500\text{m}$), and the measurement error is independent to the drifting estimates obtained with the vision-only process; the error does not accumulate over time. Moreover, it is easy to implement this approach, since various hardware systems are available for a wireless radio signal network. It should be noted that the visual-inertial system (VINS) represents the state-of-art in terms of performance in scale estimation and drift mitigation. Our approach can be either combined with the VINS to improve the robustness, or be an alternative tool.

The followings are the main contributions of the proposed sensor fusion approach:

- 1) The global scale ambiguity in monocular Vision-based SLAM (VSLAM) is reliably resolved with additional metric information.
- 2) Drifts are reduced with a fusion of visual odometry and ranging measurements without closing a loop (However, the loop closing technique can be still used).
- 3) Our approach is evaluated with real datasets collected on a rover, achieving more accurate trajectory estimates compared to the state-of-the-art visual odometry techniques.
- 4) The proposed algorithm can be simply integrated with the back-end of the monocular VSLAM methods, without significantly increasing the complexity in the systems.

We discuss related works in Section II, focusing on the relation with VINS techniques. Then, we clarify our system setup and the radio-based ranging model in Section III, followed by the algorithm overview in Section IV. In Section V, an initial global scale estimation method using ranging measurements is proposed. Then, our graph-based sensor fusion method is presented in Section VI. In Section VII, we present the evaluation results of the proposed algorithm with real datasets obtained with a rover system.

II. RELATED WORKS

A. Monocular VSLAM

A VSLAM system consists of two processes: the front-end to obtain geometric information with visual observations, and the back-end to estimate camera motions and construct a map database using the acquired information in the front-end.

In feature-based VSLAM [2] [3], feature points are extracted in the front-end process, and used for further estimation. Other approaches are available; for example, pixel intensity can be directly used to detect motion between

¹Young-Hee Lee is a research assistant at the Institute for Communications and Navigation, Technical University of Munich, Munich, Germany younghee.lee@tum.de

²Chen Zhu and Gabriele Giorgi are researchers at the Institute of Communications and Navigation, German Aerospace Center, Oberpfaffenhofen, Germany chen.zhu@dlr.de, gabriele.giorgi@dlr.de

³Christoph Günther is the head of the Institute of Communications and Navigation of the German Aerospace Center and the Institute for Communications and Navigation at the Technical University of Munich christoph.guenther@dlr.de

This work was conducted for the project VaMEx-CoSMiC that is supported by the Federal Ministry for Economic Affairs and Energy (on the basis of a decision by the German Bundestag, grant 50NA1521) and administered by DLR Space Administration.

frames under the assumption of constant light condition [4]. The latter method is more efficient compared to feature-based algorithms in terms of computational complexity, since pixels are directly used without the additional processing that is needed to extract feature information.

In the back-end of the VSLAM, camera motion and map points are estimated with the geometric information provided from the front-end. A Bayesian filter, such as an Extended Kalman Filter (EKF) is an efficient method for real-time applications [5]. Keyframe-based graph optimization is applied to carry out the estimation processes as well, such as the batch least-square estimation (bundle adjustment) [2]. To achieve real-time processing, state-of-art methods marginalize the graph, increasing the numerical efficiency. Kaess et al. propose procedural factor graph marginalization to achieve a trade-off between speed and accuracy in the VSLAM [6], [7].

Our work focuses on improving the back-end with radio-based ranging measurements. We use the graph-based framework, treating the ranging measurements as additional edges in the factor graph.

B. Visual-ranging system

Global Navigation Satellite System (GNSS) is one of the generally used positioning system, based on pseudo-range measurements derived from broadcast satellite radio signals. However, the signal from satellites is not available in many cases due to obstructions or lack of infrastructures (e.g. urban canyons, indoors, or other planets). In such environments, ranging can be provided with other systems, such as signal of opportunities (SOP) [8]. Morales et al. [9] propose navigation with a fusion of a cellular network and inertial sensors.

Tabibiazar and Basir [10] propose an approach using a monocular camera and range measurements from a cellular network, using a loosely-coupled method for localization. It requires a minimum of three ranging links, which are not available in many circumstances (e.g. mobile networks). Our method only needs a single ranging link from a radio beacon or any other SOP sources.

On-board lidar or radar sensors can provide ranging measurements as well. Zhang et al. [11] [12] propose a fusion of monocular vision and lidar-based ranging. Unlike the aforementioned methods, which use dense measurements obtained with an on-board sensor, we exploit a system that provides sparse distance information from a fixed anchor point which is independent from the camera motion estimation. Therefore, the measurement error is not accumulating over time. Furthermore, our method does not considerably increase the system complexity.

C. Visual-inertial system (VINS)

Nützi et al. [13] propose scale estimation with inertial measurements using a tightly-coupled sensor fusion. Since the VINS performs very efficiently, the majority of the state-of-art monocular VSLAM algorithms provides the VINS, integrated with their vision-only solution [14] [15] [16].

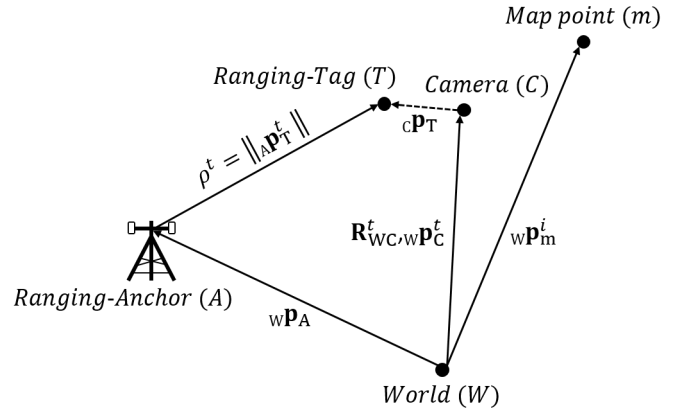


Fig. 1: System setup and the relative geometry between the frames

Despite of its verified performance, the VINS approaches have drawbacks. First, the global consistency of estimation cannot be refined since on-board inertial sensors are used in the system. Furthermore, inertial bias calibration is coupled with the VSLAM estimation, causing difficulty in error analysis and system monitoring.

We use a ranging system that provides absolute distance measurements in which errors are not coupled with the dead-reckoning vision-only estimation, hence errors can be analyzed easily. In addition, the calibration process is simpler, compared to the VINS. Furthermore, our method can be easily integrated with the VINS.

III. SYSTEM MODEL

In this section, we introduce our system setup, including the frame definition and notations that are used throughout this paper, followed by the measurement model of radio-based ranging (Two-way Time-of-Flight).

A. Frame definition

Fig. 1 shows our system setup and the relative geometry between frames. We mount a camera on a rover to log monocular image data. The *Camera frame* is attached to the current camera body frame (C), and the *World frame* (W) is defined by the initial *Camera frame* when monocular visual odometry successfully initializes map points (m). For the ranging system, we mount a ranging-tag (T) on the rover with the camera, and install a ranging-anchor (A) at a fixed position. These ranging nodes provide the relative distance measurement ρ^t between the rover and the anchor point. The position vector of the anchor with respect to the *World frame* $w\mathbf{p}_A$ can be estimated by trilaterating multiple distance measurements collected at the initial phase of a mission [17] [18]. An orthonormal rotation matrix \mathbf{R} describes a 3D camera rotation, and a 3D spatial position vector is denoted with \mathbf{p} . We use subscripts for the frame information and superscripts for the time stamps of the camera poses or map point numbers. For example, the rotation matrix from the *Camera frame* to the *World frame* at time t is \mathbf{R}_{WC}^t , and the 3D position of the i -th map point with respect to the *World frame* is $w\mathbf{p}_m^i$.

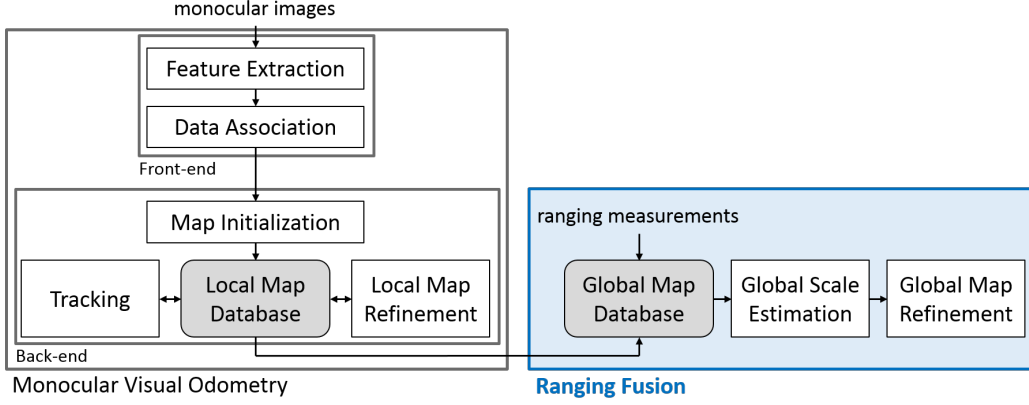


Fig. 2: The flowchart of the proposed algorithm

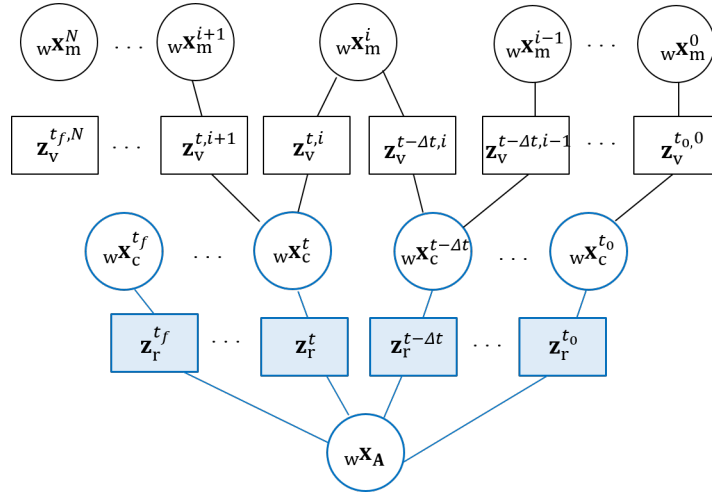


Fig. 3: The graph representation of the global map database. \mathbf{x} denotes the state variables in the graph: the camera pose with respect to the *World frame* at t ($w\mathbf{x}_c^t$), and the i -th map point position with respect to the *World frame* ($w\mathbf{x}_m^i$). Constraints are denoted with \mathbf{z} : the re-projection error of the i -th map point in the camera frame at t ($\mathbf{z}_v^{t,i}$), and the distance error at t (\mathbf{z}_r^t)

B. Radio-based ranging

We use the Two-way Time-of-Flight (TW-ToF) ranging technique [19] to obtain the relative distance measurements between two wireless sensor nodes (ranging-tag and ranging-anchor in Fig. 1). First, one of the nodes (node A) transmits a signal to the other (node B), and the node B sends the signal back to the node A. Node A collects the round-trip-time of the signal multiple times, and computes an average of the whole set. Then, we can compute the relative distance between two nodes using the averaged time \hat{t}_{ToF} :

$$d = \frac{c}{2} (\hat{t}_{ToF} - t_{off}) \quad (1)$$

where c is the speed of light, and t_{off} is a time offset including all delays in the system. Since the TW-ToF ranging system uses the signal round-trip time, it does not require time synchronization of the nodes with high precision which is problematic for other ranging systems using wireless sensor networks.

IV. ALGORITHM OVERVIEW

The flowchart of the proposed algorithm is given in Fig. 2. We use the state-of-art feature-based algorithm ORB-SLAM [3] without closing loops to perform monocular visual odometry; any other algorithms can be applied as well. First, we extract feature information (the features' positions and corresponding descriptors) from the received images and store it in a map database. Then, we try to find feature matches between the images taken at different locations (the front-end of visual odometry). If the number of matched features is larger than a given threshold, we initialize map points by triangulation. Once the map points are successfully initialized, we start to track the current features in the database to iteratively estimate camera poses. Meanwhile, a local map (a subset of the global map that only includes the camera poses and map points correlated with the current ones) is continuously optimized such that the summation of the squared re-projection errors is minimized (the back-end of visual odometry). This is an effective approach for real-time localization since it does not require

large data processing. Local map refinement is conducted in a separated processing unit, hence real-time tracking is not hindered. However, a global scale of the map cannot be estimated with monocular vision. Moreover, estimation errors are accumulating over time.

Therefore, we propose to fuse the distance measurements between a rover and an anchor point with monocular vision. First, we construct a graph-based global map database including the ranging measurements as shown in Fig. 3. The variables denoted with circles are the state variables in the graph: the keyframe poses ${}^w\mathbf{x}_C^t$ and map point positions ${}^w\mathbf{x}_M^t$. Constraints are denoted with the squares on the edges between the corresponding nodes; re-projection errors (visual constraints) $\mathbf{z}_V^{t,i}$ and distance errors (ranging constraints) \mathbf{z}_r^t . After constructing the global map, we estimate an initial guess of the global scale using the monocular visual odometry results and ranging measurements. Then, we scale the keyframe and map point positions with this value. Lastly, we perform a global least-squares estimation (global LSE) with the Levenberg-Marquardt algorithm aiming at minimizing the squared weighted sum of both re-projection and ranging residuals. This is described in the following section.

V. GLOBAL SCALE ESTIMATION

Since no metric information is available in a monocular vision-only method, the scales of the camera translation and map points are randomly determined at the initialization stage (e.g. normalization). After the initialization, the system generates the trajectory and new map points with the scales that are relative to the initialized value; if no error is introduced in the visual odometry process, we can recover the up-to-scale estimates by multiplying with a true global scale α .

To resolve this global scale ambiguity, we propose to use an additional ranging system which provides the distance measurements between a rover and an anchor $\hat{\rho}^t$. The ranging measurement model is

$$\hat{\rho}^t = \|\mathbf{A}\mathbf{p}_T^t\| = \|\mathbf{-w}\mathbf{p}_A + \alpha\mathbf{w}\mathbf{p}_C^t + \mathbf{R}_{wCC}^t\mathbf{p}_T\| \quad (2)$$

For each value of ρ , two solutions for the global scale can be found analytically, if no error is modeled for both odometry and ranging:

$$\alpha_0, \alpha_1 = -\mathbf{B}/\mathbf{A} \pm \sqrt{(\mathbf{B}/\mathbf{A})^2 - \mathbf{C}/\mathbf{A}} \quad (3)$$

$$\text{where } \begin{cases} \mathbf{A} = \|\mathbf{w}\mathbf{p}_C^t\|^2 \\ \mathbf{B} = \mathbf{w}\mathbf{p}_C^{t,T}(\mathbf{R}_{wCC}^t\mathbf{p}_T - \mathbf{w}\mathbf{p}_A) \\ \mathbf{C} = \|\mathbf{R}_{wCC}^t\mathbf{p}_T - \mathbf{w}\mathbf{p}_A\|^2 - (\hat{\rho}^t)^2 \end{cases}$$

Fig. 4 shows the physical meaning of the solutions. A single ranging measurement ρ does not allow to discriminate between two candidates (α_0, α_1). When another relative distance ρ' is available, we can compute an additional duplet. If there is no error in visual odometry and ranging measurements, one of the factors remains constant for each ranging: this value represents the recovered global scale (α_1 in Fig. 4). Moreover, the relative geometry of the estimates

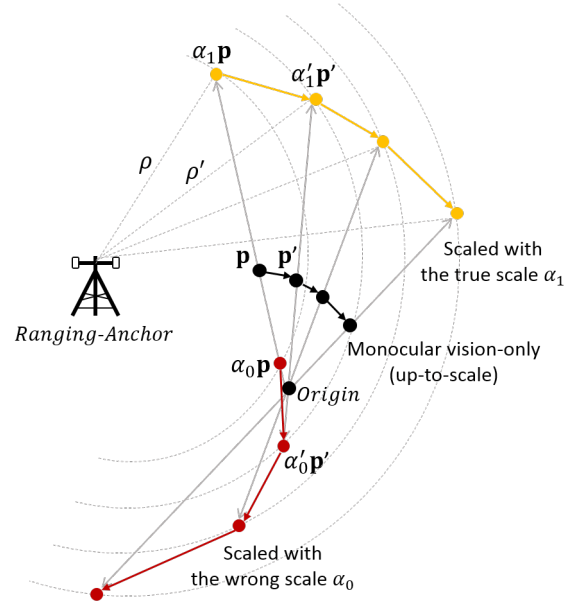


Fig. 4: When there is no error in the distance measurements ρ and the trajectory generated with a monocular vision-only method (black), the true scale α_1 is constant, enabling the recovery of the true trajectory (orange). The relative scales and direction in the up-to-scale trajectory are maintained in the true trajectory. The wrong candidate α_0 is a different value for each ranging measurement, and the relative geometry is not preserved in the trajectory scaled with the wrong value (dark red)

generated with a monocular vision-only method (black in Fig. 4) is maintained in the trajectory scaled with the true global scale (orange in Fig. 4). However, the other candidate is a different value at each time. In addition, the relative scales and direction of the up-to-scale trajectory are not maintained in the trajectory scaled with the wrong value (dark red in Fig. 4).

In the real world, errors exist in the vision-only estimates and ranging. Therefore, we collect a number of estimated duplets over time, and select the mean value of α_i characterized by the smallest standard deviation.

VI. GLOBAL MAP REFINEMENT

After the global scale initialization, all keyframes and map point positions are accordingly scaled. Then, we carry out the global least-squares estimation such that the summation of the squared re-projection and distance errors in the global map database is minimized. Ranging measurements provide absolute metric information, which is not coupled with the parameters estimated with the vision-based method, enabling mitigation of drifts in the dead-reckoning process.

VII. EXPERIMENTAL TESTS

We conducted a number of tests in an outdoor environment to collect datasets; no public available dataset provides images and ranging simultaneously. The rover system developed by the Institute of Communications and Navigation,

German Aerospace Center (DLR-KN) was used as a platform.

In this section, we present the test results of the proposed algorithm, conducted with the collected real dataset.

First, we show the initial global scale estimation results. Then, we discuss the trajectory and map points estimated with the proposed approach. Moreover, the performance of trajectory estimation is compared quantitatively with those provided by other algorithms, such as the trajectory estimated with monocular visual odometry scaled with the ground truth and the initial global scale guess, and stereo visual odometry.

A. Setup

We used the left camera of the stereo camera Bumblebee2 (FLIR) to obtain monocular images (1024×768 , 30 fps), and two ultra-wideband (UWB) wireless sensors (Decawave) to collect distance interruptions. The UWB sensors provide the distance measurements between the rover and an anchor point with an accuracy of ± 10 cm under ideal circumstances, such as no multipath or Line-of-Sight interruptions. These sensors were integrated them in the rover system developed by the Institute of Communications and Navigation, German Aerospace Center (DLR-KN); we mounted the camera and a UWB sensor (ranging-tag) on a rover, and installed another UWB sensor (ranging-anchor) at the base station. The measurement campaign was conducted on the soccer field located in the German Aerospace Center, Oberpfaffenhofen, collecting images and ranging, as well as the ground truth trajectory with a GNSS- and Inertial Measurement Unit (IMU)-based Real Time Kinetic (RTK) system for a quantitative evaluation. Fig. 5 is an example image taken with features extracted in the frame, and Fig. 6 shows the absolute differences between the ranging measurements and the true values computed with the ground truth trajectory.

B. Results

Table I shows the mean and standard deviation of two global scale candidates α_0 and α_1 . The mean value of α_1 is chosen as an initial global scale guess since the standard deviation of α_1 is smaller than the one of α_0 . Fig. 7(a) presents the trajectory scaled with α_0 (dark red) and α_1 (orange), as well as the ground truth trajectory (black) with respect to the *World frame*. It qualitatively shows that α_1 is a proper initial guess of the global scale compared to α_0 .

Fig. 8 shows the trajectory and map point estimates with respect to the *World frame* after conducting the proposed global least-squares estimation. Map point positions cannot be evaluated quantitatively since there is no ground truth, whereas the performance of the trajectory estimation can be quantified by RMSE using the ground truth obtained with the RTK system. As shown in Table II, even though we scale the monocular visual odometry trajectory with the true scales computed with the ground truth trajectory, the RMSE is relatively high. In addition, the trajectory scaled with the initial global scale guess shows considerable RMSE as well. After the global least-squares estimation, the RMSE of the trajectory estimation becomes significantly lower than the

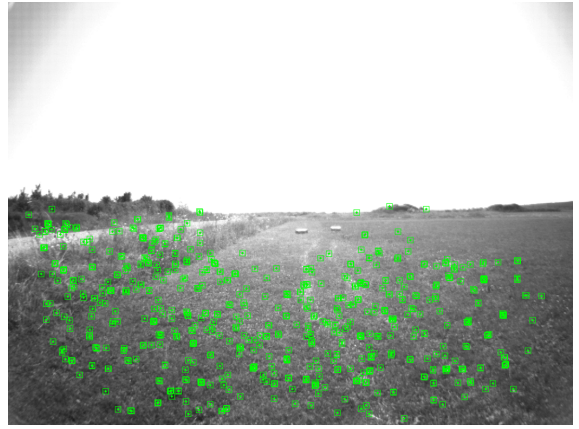


Fig. 5: An example of the image with the extracted features

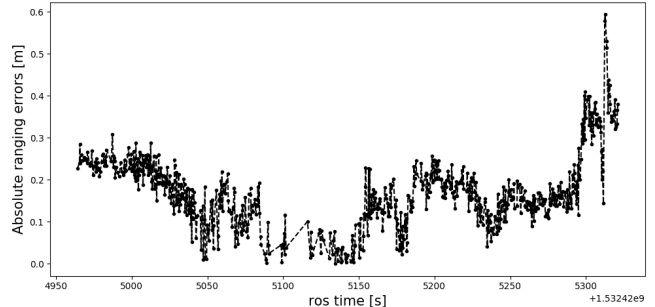


Fig. 6: Absolute differences between the ranging measurements and the true distances computed with the ground truth trajectory in [m]

others. This result shows that our global map refinement reduces accumulated estimation errors.

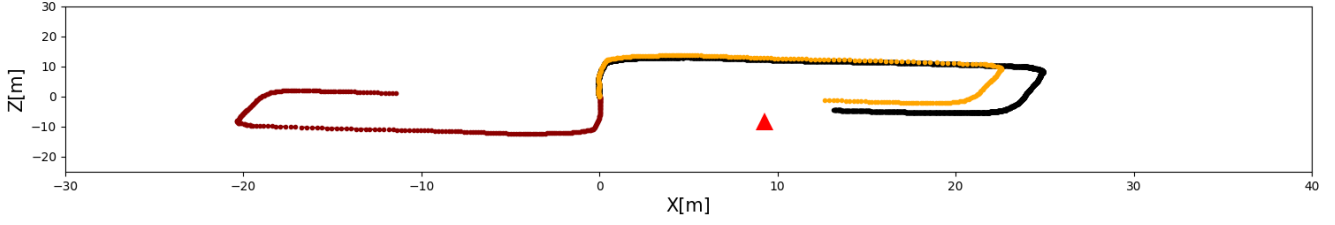
Additionally, we tested stereo ORB-SLAM [3] without closing loops, using the real stereo images collected with the stereo camera. As shown in Fig. 5, the detected features are mostly in the grass field, which cannot be easily tracked between frames (low quality features). Moreover, we used a stereo camera with a short baseline length (~ 12 cm). Therefore, the performance is worse than the one of the proposed algorithm, as well as monocular visual odometry scaled with true scales as shown in Fig. 7(c) and Table II.

VIII. CONCLUSION

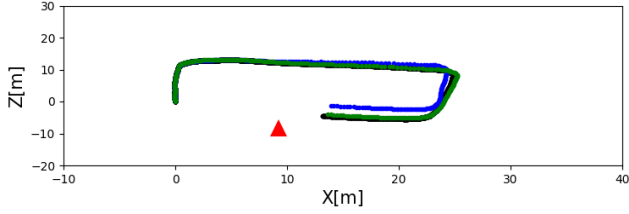
We proposed a fusion of radio-based ranging measurements and monocular vision, enabling global scale estimation and drift mitigation in monocular visual odometry with a single ranging link. We tested the algorithm with a real dataset collected on a rover, showing more accurate performance in the trajectory estimation compared to other algorithms (monocular visual odometry scaled with the true scales and the initial global scale guess, and stereo visual odometry).

The proposed method can be easily integrated within the VSLAM's back-end with a simple additional setup. In addition, our method can be an auxiliary option to the VINS (or integrated with the VINS) for resolving the scale ambiguity problem.

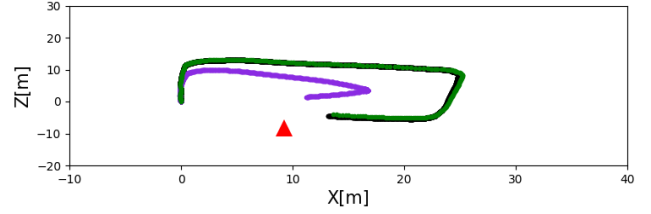
As a further step, we will test the proposed algorithm



(a) The trajectory estimates multiplied with two global scale candidates α_0 (dark red) and α_1 (orange) in Table I



(b) The trajectory obtained with the proposed algorithm (green), and the one scaled with the true scales computed with the ground truth trajectory (blue)



(c) The trajectory obtained with the proposed algorithm (green), and the one estimated with stereo visual odometry (purple)

Fig. 7: The trajectory estimates in the horizontal plane: the ground truth trajectory is represented as a black line, and the anchor position is denoted with a red triangle in each figure

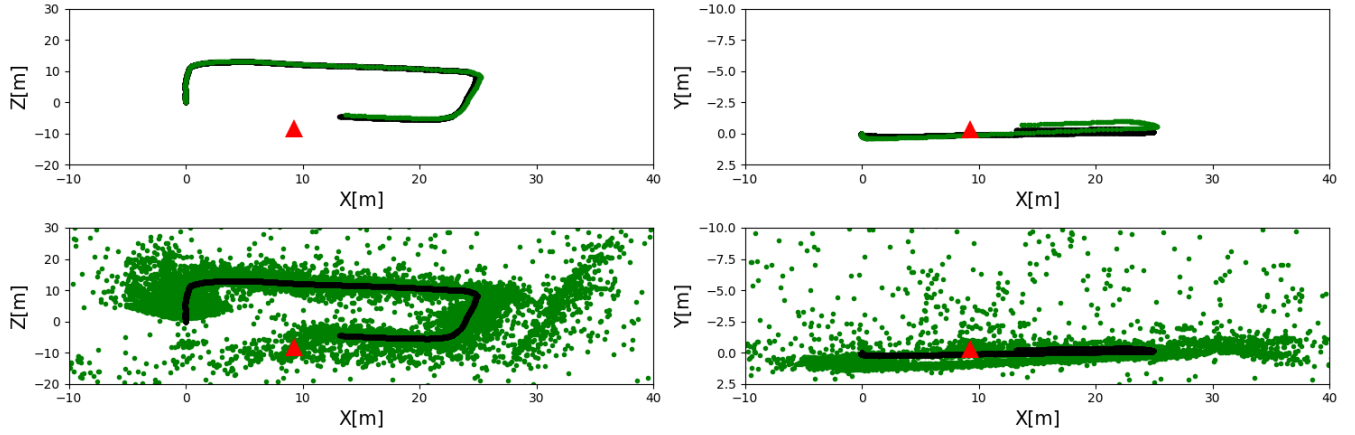


Fig. 8: The trajectory (top) and map points (bottom) obtained with the proposed fusion algorithm in the horizontal (left) and vertical (right) planes

TABLE I: The mean and standard deviation of two global scale candidates

	Mean	Standard deviation
α_0	-4.17613	9.39333
α_1	4.63161	0.213666

TABLE II: The RMSE of the trajectory estimates in [m]

Algorithms	RMSE[m]
MonoVO scaled with the true scales (blue in Fig. 7)	1.742
MonoVO scaled with the estimated global scale, without global LSE (orange in Fig. 7)	2.206
Trajectory estimated with the proposed approach, with global LSE (green in Fig. 7)	0.450
StereoVO (purple in Fig. 7)	6.833

on a number of different scenarios. The initial global scale estimation method will be improved as well, improving robustness against outliers in ranging measurements. Furthermore, real-time performance will be improved, implementing the algorithm on an additional processing unit (separated

from real-time tracking).

REFERENCES

- [1] 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.

- [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] 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.
- [4] 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.
- [5] A. Davison, I. Reid, N. Molton, and O. Stasse, "Monoslam: Real-time single camera slam," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 29, no. 6, pp. 1052–1067, June, 2007.
- [6] M. Kaess, "Incremental Smoothing and Mapping," *Work*, no. December, 2007.
- [7] M. Kaess, H. Johannsson, R. Roberts, V. Ila, J. J. Leonard, and F. Dellaert, "iSAM2: Incremental smoothing and mapping using the Bayes tree," *The International Journal of Robotics Research*, vol. 31, no. 2, pp. 216–235, 2012. [Online]. Available: <http://journals.sagepub.com/doi/10.1177/0278364911430419>
- [8] J. Raquet and R. K. Martin, "Non-gnss radio frequency navigation," in *2008 IEEE International Conference on Acoustics, Speech and Signal Processing*, March 2008, pp. 5308–5311.
- [9] J. Morales, P. Roysdon, and Z. Kassas, "Signals of opportunity aided inertial navigation," in *Proceedings of ION GNSS Conference*, 2016, pp. 1492–1501.
- [10] 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.
- [11] J. Zhang and S. Singh, "Visual-lidar odometry and mapping: Low-drift, robust, and fast," *IEEE Intl. Conf. on Robotics and Automation (ICRA)*, May 2015.
- [12] J. Zhang, M. Kaess, and S. Singh, "A real-time method for depth enhanced visual odometry," *Auton. Robots*, vol. 41, no. 1, pp. 31–43, Jan. 2017. [Online]. Available: <https://doi.org/10.1007/s10514-015-9525-1>
- [13] 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.
- [14] R. Mur-Artal and J. D. Tards, "Visual-inertial monocular slam with map reuse," *IEEE Robotics and Automation Letters*, vol. 2, no. 2, pp. 796–803, April 2017.
- [15] V. Usenko, J. Engel, J. Steckler, and D. Cremers, "Direct visual-inertial odometry with stereo cameras," in *2016 IEEE International Conference on Robotics and Automation (ICRA)*, May 2016, pp. 1885–1892.
- [16] L. von Stumberg, V. Usenko, and D. Cremers, "Direct sparse visual-inertial odometry using dynamic marginalization," in *icra*, May 2018.
- [17] F. Thomas and L. Ros, "Revisiting trilateration for robot localization," *IEEE Transactions on robotics*, vol. 21, no. 1, pp. 93–101, 2005.
- [18] C. Zhu, G. Giorgi, and C. Günther, "Planar pose estimation using a camera and single-station ranging measurements," in *Proceedings ION GNSS+ 2017, Portland, Oregon*, 2017.
- [19] E. B. Mazomenos, D. De Jager, J. S. Reeve, and N. M. White, "A two-way time of flight ranging scheme for wireless sensor networks," in *European Conference on Wireless Sensor Networks*. Springer, 2011, pp. 163–178.