2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)
Daejeon Convention Center
October 9-14, 2016, Daejeon, Korea

Stereo Vision-based Localization for Hexapod Walking Robots Operating in Rough Terrains

Thomas Fischer[1] Taihú Pire [1] Petr Čížek[2] Pablo De Cristóforis[1] and Jan Faigl[2], *Member, IEEE*

***Abstract*— This paper concerns the self-localization problem of a hexapod walking robot operating in rough terrains. Given that legged robots exhibit higher terrain passability than wheeled or tracked platforms when operating in harsh environments, they constitute a challenge for the localization techniques because the camera motion between consecutive frames can be arbitrary due to the motion gait and terrain irregularities. In this paper, we present and evaluate an inertially assisted Stereo Parallel Tracking and Mapping (S-PTAM) method deployed on a hexapod crawling robot in a rough terrain. The considered deployment scenario is motivated by autonomous navigation in an unknown environment in an open loop fashion. The reported results and comparison with an existing RGB-D SLAM technique show the feasibility of the proposed approach and its suitability for navigation of crawlers in harsh environments.**

I. Introduction

Legged robots have attracted an attention from the robotics community in recent years due to their ability to work in harsh environments inaccessible for wheeled robots while being able to carry more load and operate longer than Unmanned Aerial Vehicles (UAV). Their skills to perform omnidirectional motion and traverse rough terrains make them a great choice for many practical robotic applications, such as Urban Search and Rescue (USAR) missions in collapsed buildings or disaster areas. In such scenarios, the walking robot should have enough autonomy to take advantage of its high locomotion capabilities. Without a reliable external localization (like GPS or motion capture system), it is necessary that the autonomous mobile robot builds a spatial representation of the environment (map) and localizes itself within it. This problem is referred to as Simultaneous Localization and Mapping (SLAM) [1] in the robotics community. An efficient and reliable SLAM method is the core part of an autonomous navigation system, which provides both the pose information and the model of the environment. This information is in turn indispensable for the efficient motion planning and control of a walking robot.

This work is a direct result of the bilateral cooperation program between the Czech and Argentinian Republics support by the Argentinian project ARC/14/06 and travel support of the Czech Ministry of Education under the project No. 7AMB15AR029. The work of J. Faigl and P. Čížek has been also supported by the Czech Science Foundation (GAČR) under research project No. 15-09600Y.

[1]Thomas Fischer, Taihú Pire and Pablo De Cristóforis are with the Computer Science Department, University of Buenos Aires, Argentina. {tfischer, tpire, pdecris}@dc.uba.ar

[2]Petr Čížek and Jan Faigl are with the Department of Computer Science, Faculty of Electrical Engineering, Czech Technical University in Prague, Technická 2, 166 27, Prague, Czech Republic. {petr.cizek, faiglj}@fel.cvut.cz

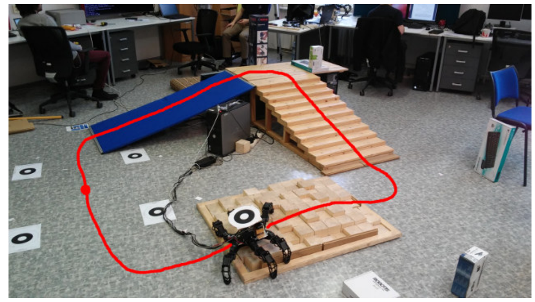

Fig. 1. Experimental set-up.

Legged robots moving on uneven ground require 3D positioning with regard to six degrees of freedom. A 6 DoF pose contains the position of the robot in three dimensions and its orientation as the yaw, pitch, and roll angles. Some early works made used of legged odometry methods to determine the pose of the robot [2]. However, legged odometry often returns erroneous pose estimates due to the presence of slipping surfaces and mechanical imperfections. In real USAR missions, a precise self-localization capability based on exteroceptive sensing is therefore required. In this context, visual sensors such as cameras or structured light range sensors (SLRS) become attractive choice to face the SLAM problem. While being affordable, small and light, they can provide high resolution 3D data of the environment in real time with low power consumption.

Visual SLAM has become one of the most studied topics in mobile robotics in the latest two decades. Nowadays, there are robust and accurate Visual SLAM solutions that are capable of working in real-time. Different Visual SLAM techniques have been deployed to nearly all kinds of robots, e.g., wheeled [3], [4], flying [5], [6], walking [7], [8], or even underwater vehicles [9]. In contrast to wheeled and flying robots, the application of SLAM techniques to legged robots is scarce. Recently, new methods have been proposed for fusing Visual SLAM techniques with inertial sensors for field robotics applications that may also be beneficial for navigation of legged robots.

In this work, we focus on the deployment of the recently released Stereo Parallel Tracking and Mapping method (called S-PTAM) [10] on a hexapod crawling platform operating in rough terrain environment with different obstacles, e.g., see Fig. 1. In such a scenario, a precise estimation of the full robot 6 DoF pose is required under presence of unpredictable camera shaking and motion blur induced by the robot locomotion. Hence, we propose to extend the original S-PTAM method to consider additional information from the inertial measurement unit (IMU). The deployment scenario is motivated by autonomous navigation in unknown environment, and therefore, we do not consider loop closure, which may significantly improve the localization precision, but we rather evaluate the method in an open loop fashion. A comparison with an existing RGB-D SLAM [11] technique that uses SLRS (Kinect-style camera) is presented to evaluate suitability of the proposed approach.

The paper is organized as follows. Related work on the legged robot localization is presented in Section II. Section III provides description of the proposed IMU assisted S-PTAM method and a brief description of the utilized RGB-D SLAM is presented in Section III-B. Evaluation results of experimental validation with real hexapod robot are reported in Section IV. Concluding remarks are dedicated to Section V.

II. Related Work

This section provides a brief overview of the recent work addressing the problem of how walking gaits affect vision-based SLAM techniques.

In [12], a robot motion model based on the control commands is proposed to improve the MonoSLAM system of Davison [7] over legged robots. Although the accuracy of the SLAM system improved with the proposed model, the frame rate decreased to 2 Hz, which is in a huge contrast to the original MonoSLAM that works up to 30 Hz.

Chilian et al. [13] propose a navigation system for a legged robot crawling on rough terrains based solely on stereo vision. A visual SLAM system is used for the localization of the robot and construction of the terrain map containing the traversability information. A D* Lite planner is used to guide the robot through the terrain to a predefined goal. The work reports that an error of reaching the goal location is less than 3% of the traveled path length about few meters long. The whole system operates at 1 Hz, which is far to be useful for real-time applications. An improved version of the system by considering measurements provided by legs odometry and IMU has been presented in [14]; however, the frame rate of the system was not addressed.

Stelzer et al. [15] developed a complete visual stereo-based navigation framework for their hexapod robot. Pose estimates are obtained by fusing inertial data with relative leg and visual odometry measurements using an indirect information filter. In a similar way, [16] fuses the information from stereo vision, leg odometry, and IMU to obtain accurate state estimate of the quadruped robot.

Cited works used the classical EKF-based approaches to face the SLAM problem (EKF-SLAM). A comparison between filter-based Visual SLAM approaches and keyframe-based methods which use global optimization (i.e., bundle adjustment) throws the conclusion that the latter ones are found to achieve the best balance between precision and computational cost [17].

One of the most actively developed keyframe-based approaches is Parallel Tracking and Mapping (PTAM) [18] that separates the mapping process from camera tracking in two different threads, performing an ordinary batch estimation on a small number of keyframes as one process while tracking the camera relative position to the map as another process. The recently proposed stereo version of the PTAM (called S-PTAM) [10] allows to reconstruct a metric 3D map for each frame, improve the accuracy of the mapping process with respect to the monocular PTAM, and it avoids the well-known bootstrapping problem. Besides, the stereo approach allows to compute the real scale of the environment, which is an essential feature for robots that have to interact with their surrounding workspace. In this paper, we consider the recent S-PTAM system as a suitable localization and mapping approach for hexapod walking robots operating in rough environments because of its performance and accuracy regarding the Visual SLAM state-of-the-art [19]. Moreover, it is available as an open-source project.[1]

III. SLAM Methods

A. S-PTAM

The feature-based S-PTAM system [10] is a stereo Visual SLAM method designed for localization and mapping in large scale environments. S-PTAM is based on the monocular Parallel Tracking and Mapping (PTAM) method introduced in [18]. The method consists of two processes working in parallel. The tracking part, which tracks the camera pose for each pair of incoming images, and the mapping part, which builds and refines a map of selected 3D features.

When tracking, for each new pair of stereo images, visual features are extracted using GFTT [20] and BRIEF [21] descriptors. The map points are then projected onto the image plane using a prediction of the current camera pose, which in turn are matched to the closest image features in the region of interest. The size of the search region determines the computational load of the matching process. Note that the tracking phase should be fast enough to allow real-time responses, since the better the prediction, the smaller this search region can be. S-PTAM uses a decay velocity model to compute the predicted pose.

The matches can be then used to refine the estimated camera pose using an iterative least squares minimization method, e.g., using the Levenberg-Marquardt algorithm. Finally, stereo matches are computed between left and right image features that could not be matched to the map, and are in turn triangulated and then inserted as new map points.

Concurrently, a map refinement process is running, which is also based on the Levenberg-Marquardt algorithm. This continuously performs Bundle Adjustment optimization on the current local portion of the map determined by a subset of the last salient frames called keyframes. Fig. 2 illustrates the general overview of the S-PTAM system.

As mentioned in [10], S-PTAM uses a decay velocity model to generate a pose prediction for each incoming stereo frame. This model assumes the motion transitions are sufficiently smooth which is mostly satisfied for wheeled robots, while not for hexapod crawlers operating in rough

[1] https://github.com/lrse/sptam

terrains. Therefore, in this work, the original model is replaced by an Extended Kalman Filter (EKF), which is fed with IMU measurements (angular velocity and linear acceleration) and refined pose updates estimated by S-PTAM. This is known in literature as loosely coupled approach to sensor integration. The EKF state models position, linear velocity and acceleration, orientation and angular velocity, all having three degrees of freedom.

Loose coupling is suboptimal in the sense that the existing cross-correlations between internal states of different devices are discarded. However, we show that the original S-PTAM system [10] can be significantly improved using a loosely coupled fusion with IMU measurements. It also eases the integration of other sources of displacement information, i.e., odometry. Fig. 3 illustrates how IMU information is integrated into S-PTAM through the Extended Kalman Filter.

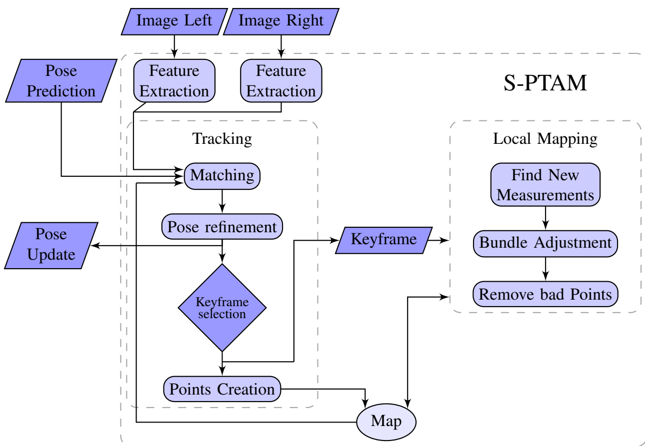

Fig. 2. Overview of the S-PTAM system.

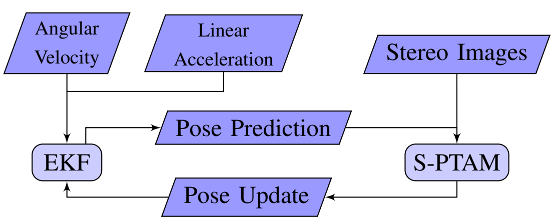

Fig. 3. S-PTAM with IMU loosely coupled integration through an EKF.

B. RGB-D Method

The RGB-D method chosen for a comparison to the Vision-based stereo SLAM is the RGB-D SLAM based on [11]. It benefits from integrating the scale information provided by 3D depth sensing into a visual SLAM system to provide a reliable localization for the legged robot. An overview of the system is depicted in Figure 4 and it operates as follows.

First, salient image points are extracted from the RGB image and descriptors are computed using the SURF algorithm [22]. Then, 3D position of each feature is obtained by extracting the depth information from the depth image at the coordinates of the feature. Since the depth information

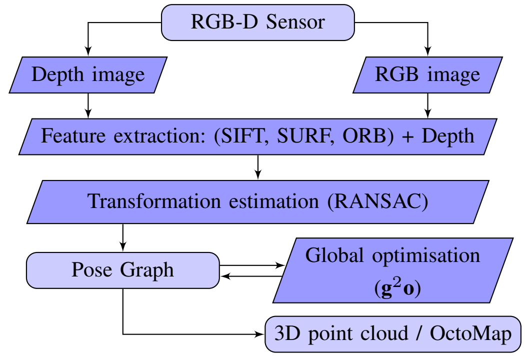

Fig. 4. Overview of the RGB-D SLAM system.

provided by the structured light sensor might not be available in certain portions of the image (e.g., due to the sunlight, maximum and minimum distance or occlusion near the edges of obstacles) features without depth information are directly discarded at this stage. Next, pairwise correspondences between the current image features and a set of the mapped environment points are computed. This set consists of directly n_p preceding frames, n_g graph neighboring frames and n_r random frames from the whole trajectory which are used for discovering of large loop closures. An estimation of the rigid transformation is computed between successfully established frame-to-frame correspondences using the RANSAC algorithm [23]. Then, the frame is added as a node to the pose graph (map) of the SLAM method. The node contains the estimated 6-DOF pose of the frame together with the set of the detected image features and the estimated dense point cloud given by the depth measurement. The edges of the pose graph represent the pairwise rigid transformations between the individual frames.

The pairwise transformations between RGB-D sensor poses in the pose graph are optimized using the $\mathbf{g^2o}$ graph optimisation framework [24] to further refine the map and provide a reliable localization. The $\mathbf{g^2o}$ framework provides a globally consistent trajectory estimation which is especially beneficial in loop-closures, when the robot revisits some previously visited area that is mapped in the constructed pose graph representation of the operational environment.

Based on the estimated trajectory resulting from the pose graph, an environment map is built by projecting the sensor depth measurements directly in a form of point clouds or voxelized OctoMap [25].

The whole method evaluated in this paper is implemented using the ROS framework [26]. The considered parametrization of the RGB-D SLAM used in the evaluation has been chosen with respect to a thorough evaluation [27]. The parametrization is the SURF feature extractor detecting 500 features and keeping 250 best matches. The comparison horizons were set to $n_p = 3, n_g = 3$ and $n_r = 0$ as we want to disable the loop-closure detection.

IV. Experiments

The experimental evaluation was performed on a laboratory test-track simulating rough terrain conditions. The experimental setup is depicted in Fig. 1. The test track consists of a square path of approximately 9 m length involving a set of stairs, a ramp, and a hill of irregular-height cubes. Each of these terrain types represents a different challenge for the visual SLAM system. The pile of irregular wooden blocks simulates an uneven terrain where the SLAM method needs to deal with the constant unpredictable shaking of the camera. While going up the stairs the robot motion is subjected to slippage, which causes the camera to change its pose abruptly. During the ramp descent, the camera is pointing to the ground, which usually has low texture; so, a limited number of features are available. The in-place turns also represent a challenge for the visual SLAM system due to the significant motion blur.

The utilized hexapod crawler is based on the *PhantomX Mark II* platform. The robot control is based on the adaptive motion gait [28], specifically designed for crawling rough terrains. The robotic platform is further equipped with the *Bumblebee* 2 stereo camera, *Asus Xtion Pro Live* RGB-D sensor, and the *XSens MTi-30* inertial measurement unit. The robot with the sensors is shown in Fig. 5.

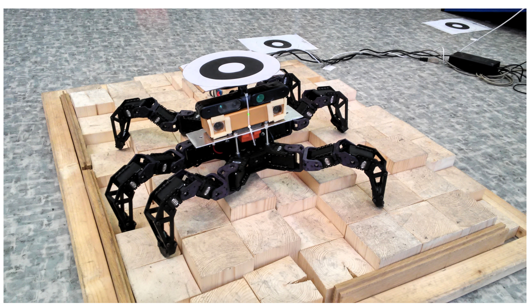

Fig. 5. The used hexapod walking robot.

The robot has been guided along a predefined path by an operator to record the required sensory data. Both SLAM algorithms were then executed offline, but inside the ROS [26] framework to simulate real time behavior. Unfortunately, the camera and RGB-D data could not be recorded simultaneously, because the infrared pattern projected by the RGB-D sensor interferes with the image sensor of the *Bumblebee* camera, causing a constant noise pattern. Therefore, several runs of the same experiment were performed for each SLAM system and the errors were aggregated over all of them to compensate over differences in the recordings.

The WhyCon [29] external localization system has been utilized to provide ground truth data of the performed trajectories. Position estimation errors are considered to be below 1% of the distance from the camera to the marker. For a single visual marker, the WhyCon system provides a realiable estimation only for the 2D position and precision of the 3D estimation significantly depends on the camera resolution and pattern size. During the experimenting, it has been observed that much more reliable estimation of the robot orientation is provided from the utilized attitude heading reference system (AHRS) *XSens MTi-30*. Therefore, the robot orientation estimated by the AHRS has been used instead of WhyCon data.

A. Evaluation metric

It is hard to judge the performance of a SLAM system solely based on absolute error measurements, since localization is performed relative to the computed map, and different maps may yield different geometries or biases, even when running the same SLAM algorithm. Therefore we employ a commonly used approach [30], [31] specifically designed for evaluating the performance of SLAM systems to cope with these differences. The approach works as follows.

Let $\boldsymbol{x}_k$ be the estimated pose at the frame k and $\boldsymbol{x}_k^*$ be the corresponding ground truth pose. Let us note the set of differences (or motions) between two frames of a sequence as $\boldsymbol{\delta}_{i,j} = \boldsymbol{x}_j \ominus \boldsymbol{x}_i$, where $\ominus$ is the inverse of the standard motion composition operator [32]. Analogously $\boldsymbol{\delta}_{i,j}^* = \boldsymbol{x}_j^* \ominus \boldsymbol{x}_i^*$. The *relative error* committed between the frames i and j can be then defined as $\boldsymbol{\delta}_{i,j} \ominus \boldsymbol{\delta}_{i,j}^*$.

In practice, the inverse motion composition operation between two poses can be computed from the corresponding transformation matrices representing each pose, namely $\boldsymbol{T}_{\boldsymbol{x}_i}$ and $\boldsymbol{T}_{\boldsymbol{x}_j}$ as

$$\boldsymbol{x}_i \ominus \boldsymbol{x}_j = \boldsymbol{T}_{\boldsymbol{x}_j}^{-1} \boldsymbol{T}_{\boldsymbol{x}_i} \tag{1}$$

These equations intentionally leave open the choice of which relative displacements $\boldsymbol{\delta}_{i,j}$ are included in the metric. As discussed by the original authors [30], different choices will highlight different properties of the data. In our case, we strive for local consistency, which is better highlighted by taking displacements as small as possible. Therefore, relative displacements are taken between the consecutive frames.

Moreover, to obtain meaningful numerical results, we need to separate the translational $\boldsymbol{\epsilon}_t$ and rotational $\boldsymbol{\epsilon}_\theta$ parts of this error, since they are different in nature. This separation was also suggested by the original authors [30]. Since a relative error $\boldsymbol{\epsilon}$ is a product of transformation matrices and thus, it is itself a transformation matrix $\boldsymbol{T}_{\boldsymbol{\epsilon}} = \begin{bmatrix} \mathbf{R} & \mathbf{t} \end{bmatrix}$ representing the error displacement. The translational and rotational parts can be extracted as follows:

$$\boldsymbol{\epsilon}_t = \boldsymbol{t}^T \boldsymbol{t} \tag{2}$$

$$\boldsymbol{\epsilon}_\theta = acos\left(\left(Tr\left(\boldsymbol{R}\right) - 1\right)/2\right). \tag{3}$$

B. Results

An example of trajectories estimated by the RGB-D and S-PTAM are shown together with their respective ground truths in Fig. 6. Both trajectories look like to perform relatively well and at a first sight the RGB-D seems to outperform the S-PTAM in terms of precision. However, only absolute errors are not appropriate indicators of the SLAM performance due to biases and localization relative to the computed map. Therefore, a relative error metric provides further insights to the system performance. We can observe a slightly better performance of the S-PTAM when comparing the relative error metric shown in Fig. 7. Boxes represent interquartile range (IQR), whiskers reach to $-1.5 \times IQR$ and $1.5 \times IQR$, and the points represent data beyond those ranges, considered

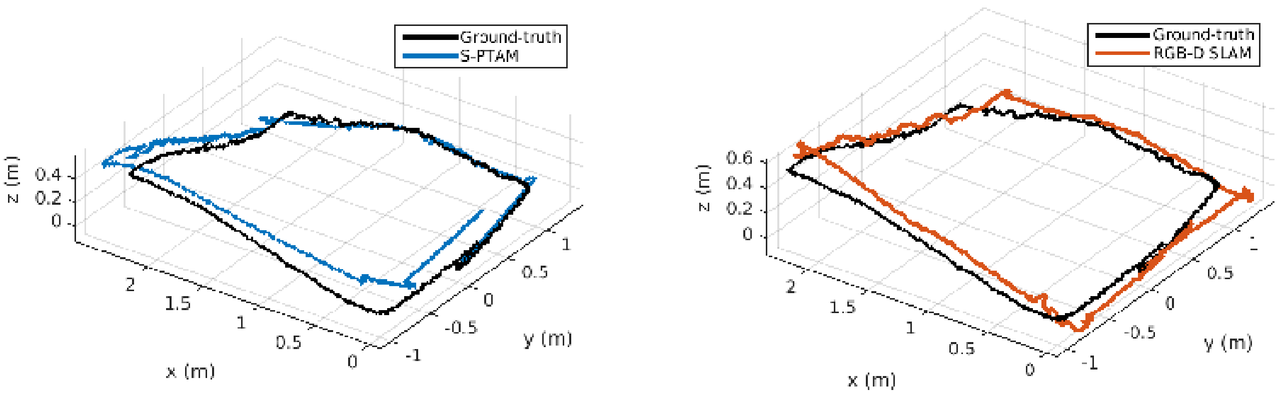

Fig. 6. RGB-D and S-PTAM trajectories with the ground truth.

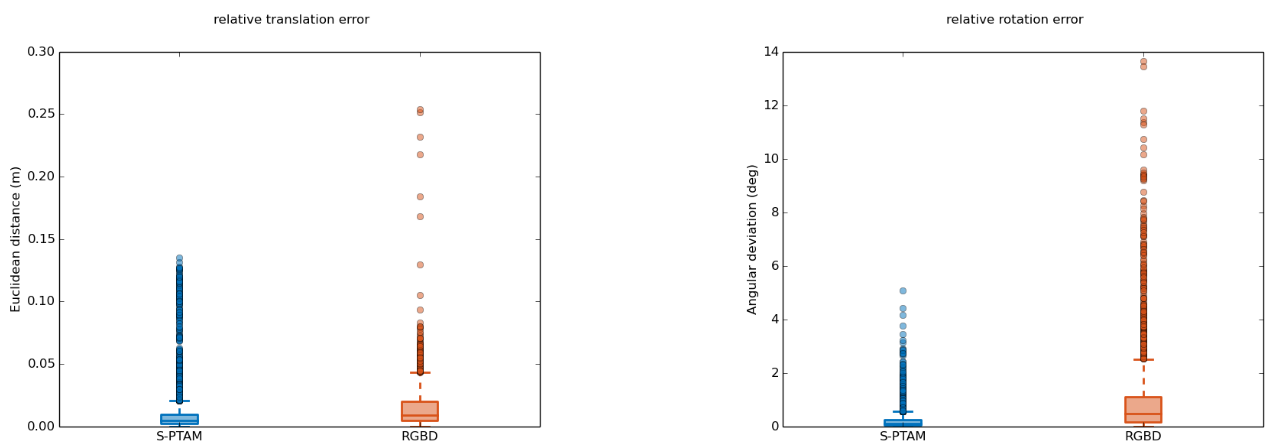

Fig. 7. Boxplots showing relative translation and rotation errors committed along the trajectory of both methods with respect to the ground truth. Measurements were aggregated over all experiments.

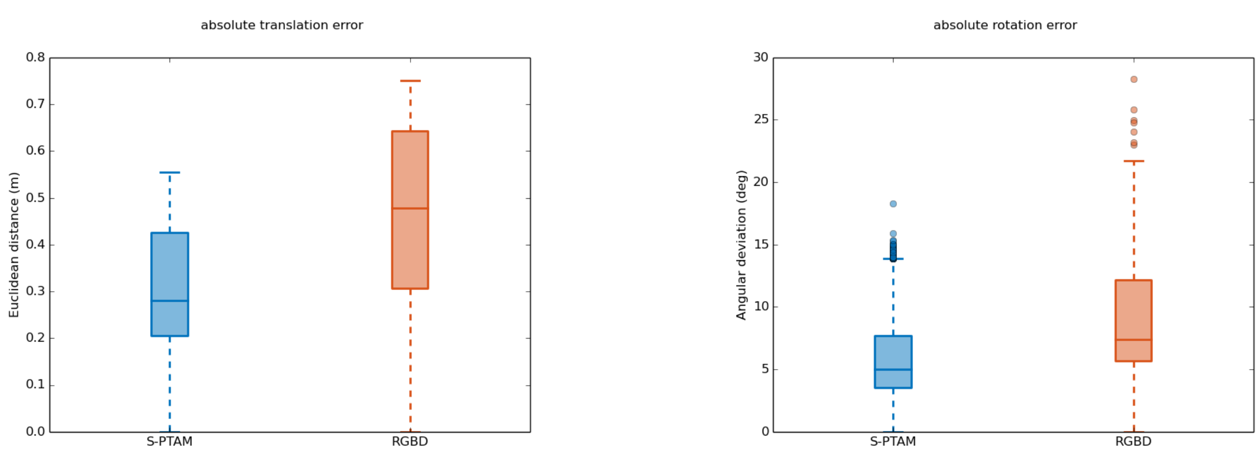

Fig. 8. Boxplots showing absolute translation and rotation errors committed along the trajectory of both methods with respect to the ground truth. Measurements were aggregated over all experiments.

outliers. The line inside the box represents the median. The absolute error is presented for completion in Fig. 8.

Regarding the real computational requirements, the S-PTAM outperforms the RGB-D SLAM by running at 10 Hz over 3-4 Hz of the RGB-D SLAM. Therefore, the proposed extension of the S-PTAM method seems to be more suitable for localization and mapping with hexapod walking robots.

V. CONCLUSIONS

In this paper, we propose an improvement of the stereo SLAM system S-PTAM, originally proposed in [10], to work on hexapod platforms crawling over rough terrains. In particular, a loosely coupled integration using an EKF filter has been considered to utilize angular velocity and linear acceleration measurements provided by inertial measurement unit. The new system has been validated in a challenging indoor environment with specifically designed obstacles, such as ramps, stairs, and irregularities. Moreover, we compare the performance of the newly proposed system with the state-of-the-art RGB-D SLAM method [11]. The results show that the proposed system is suitable for hexapod robot localization and it outperforms the RGB-D SLAM system in terms of accuracy and speed. Besides, due to limitations of the utilized RGB-D camera, the further advantage of the proposed system is ability to work in outdoor environments.

As future work, we plan to add tightly coupled integration of the IMU [33], [34] to include IMU measurements directly in the Bundle Adjustment phases of S-PTAM to increase the robustness of the whole system. In addition, we plan to deploy the SLAM method to the onboard computer of the hexapod robot to achieve autonomous navigation and perform experimental evaluation in outdoor environments.

REFERENCES

[1] S. Thrun, W. Burgard, and D. Fox, *Probabilistic robotics*. MIT press, 2005.

[2] B. Gassmann, F. Zacharias, J. M. Zollner, and R. Dillmann, "Localization of walking robots," in *ICRA*, 2005, pp. 1471–1476.

[3] K. Konolige, M. Agrawal, and J. Solà, *The 13th International Symposium Robotics Research*. Springer Berlin Heidelberg, 2011, ch. Large-Scale Visual Odometry for Rough Terrain, pp. 201–212.

[4] R. Mur-Artal, J. M. M. Montiel, and J. D. Tardós, "ORB-SLAM: a versatile and accurate monocular SLAM system," *IEEE Transactions on Robotics*, vol. 31, no. 5, pp. 1147–1163, 2015.

[5] M. Burri, H. Oleynikova, M. W. Achtelik, and R. Siegwart, "Real-time visual-inertial mapping, re-localization and planning onboard mavs in unknown environments," in *International Conference on Intelligent Robots and Systems*. IEEE, 2015, pp. 1872–1878.

[6] S. A. Scherer and A. Zell, "Efficient onbard rgbd-slam for autonomous mavs," in *International Conference on Intelligent Robots and Systems*. IEEE, 2013, pp. 1062–1068.

[7] A. J. Davison, I. D. Reid, N. D. Molton, and O. Stasse, "Monoslam: Real-time single camera slam," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 6, pp. 1052–1067, June 2007.

[8] D. Belter and P. Skrzypczynski, "Precise self-localization of a walking robot on rough terrain using parallel tracking and mapping," *Industrial Robot: An International Journal*, vol. 40, no. 3, pp. 229–237, 2013.

[9] A. Kim and R. M. Eustice, "Real-time visual slam for autonomous underwater hull inspection using visual saliency," *IEEE Transactions on Robotics*, vol. 29, no. 3, pp. 719–733, 2013.

[10] T. Pire, T. Fischer, J. Civera, P. De Cristóforis, and J. J. Berlles, "Stereo parallel tracking and mapping for robot localization," in *IROS*, 2015, pp. 1373–1378.

[11] F. Endres, J. Hess, N. Engelhard, J. Sturm, D. Cremers, and W. Burgard, "An evaluation of the RGB-D SLAM system," in *ICRA*, 2012, pp. 1691–1696.

[12] A. Schmidt and A. Kasiński, *Proc. of the International Conference Computer Vision and Graphics*. Springer Berlin Heidelberg, 2010, ch. The Visual SLAM System for a Hexapod Robot, pp. 260–267.

[13] A. Chilian and H. Hirschmüller, "Stereo camera based navigation of mobile robots on rough terrain," in *Proc. of the International Conference on Intelligent Robots and Systems*, 2009, pp. 4571–4576.

[14] A. Chilian, H. Hirschmller, and M. Görner, "Multisensor data fusion for robust pose estimation of a six-legged walking robot," in *Proceddings of the International Conference on Intelligent Robots and Systems*, 2011, pp. 2497–2504.

[15] A. Stelzer, H. Hirschmüller, and M. Görner, "Stereo-vision-based navigation of a six-legged walking robot in unknown rough terrain," *The International Journal of Robotics Research*, vol. 31, no. 4, pp. 381–402, 2012.

[16] J. Ma, S. Susca, M. Bajracharya, L. Matthies, M. Malchano, and D. Wooden, "Robust multi-sensor, day/night 6-dof pose estimation for a dynamic legged vehicle in gps-denied environments," in *ICRA*, 2012, pp. 619–626.

[17] H. Strasdat, J. Montiel, and A. J. Davison, "Real-time monocular SLAM: Why filter?" in *ICRA*, 2010, pp. 2657–2664.

[18] G. Klein and D. Murray, "Parallel Tracking and Mapping for Small AR Workspaces," in *ISMAR*. Washington, DC, USA: IEEE Computer Society, 2007, pp. 1–10.

[19] N. Krombach, D. Droeschel, and S. Behnke, "Combining Feature-based and Direct Methods for Semi-dense Real-time Stereo Visual Odometry," in *Proc. of the 14th International Conference on Intelligent Autonomous Systems*, July 2016.

[20] J. Shi and C. Tomasi, "Good features to track," in *CVPR*, Jun 1994, pp. 593–600.

[21] M. Calonder, V. Lepetit, C. Strecha, and P. Fua, "BRIEF: Binary Robust Independent Elementary Features," in *ECCV*, ser. Lecture Notes in Computer Science. Springer Berlin Heidelberg, 2010, vol. 6314, pp. 778–792.

[22] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, "Speeded-up robust features (SURF)," *Computer vision and image understanding*, vol. 110, no. 3, pp. 346–359, 2008.

[23] M. A. Fischler and R. C. Bolles, "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography," *Communications of the ACM*, vol. 24, no. 6, pp. 381–395, 1981.

[24] R. Kmmerle, G. Grisetti, H. Strasdat, K. Konolige, and W. Burgard, "G2o: A general framework for graph optimization," in *ICRA*, 2011, pp. 3607–3613.

[25] A. Hornung, K. M. Wurm, M. Bennewitz, C. Stachniss, and W. Burgard, "OctoMap: an efficient probabilistic 3D mapping framework based on octrees," *Autonomous Robots*, vol. 34, no. 3, 2013.

[26] M. Quigley, K. Conley, B. P. Gerkey, J. Faust, T. Foote, J. Leibs, R. Wheeler, and A. Y. Ng, "ROS: an open-source Robot Operating System," in *ICRA Workshop on Open Source Software*, 2009.

[27] P. Čížek and J. Faigl, "On localization and mapping with rgb-d sensor and hexapod walking robot in rough terrains," in *Proc. of IEEE SMC*, 2016.

[28] J. Mrva and J. Faigl, "Tactile sensing with servo drives feedback only for blind hexapod walking robot," in *10th International Workshop on Robot Motion and Control*, July 2015, pp. 240–245.

[29] T. Krajník, M. Nitsche, J. Faigl, P. Vaněk, M. Saska, L. Přeučil, T. Duckett, and M. Mejail, "A practical multirobot localization system," *Journal of Intelligent & Robotic Systems*, 2014.

[30] R. Kümmerle, B. Steder, C. Dornhege, M. Ruhnke, G. Grisetti, C. Stachniss, and A. Kleiner, "On measuring the accuracy of SLAM algorithms," *Autonomous Robots*, vol. 27, no. 4, pp. 387–407, 2009.

[31] A. Geiger, P. Lenz, C. Stiller, and R. Urtasun, "Vision Meets Robotics: The KITTI Dataset," *Int. J. Rob. Res.*, vol. 32, no. 11, pp. 1231–1237, September 2013.

[32] R. Smith, M. Self, and P. Cheeseman, "Estimating uncertain spatial relationships in robotics," in *ICRA*, vol. 4, 1987, pp. 850–850.

[33] S. Leutenegger, S. Lynen, M. Bosse, R. Siegwart, and P. Furgale, "Keyframe-based visual-inertial odometry using nonlinear optimization," *Int. J. Rob. Res.*, vol. 34, no. 3, pp. 314–334, Mar. 2015.

[34] M. Bloesch, S. Omari, M. Hutter, and R. Siegwart, "Robust visual inertial odometry using a direct ekf-based approach," in *IROS*, 2015, pp. 298–304.