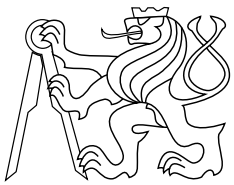
CENTER FOR
MACHINE PERCEPTION



CZECH TECHNICAL
UNIVERSITY IN PRAGUE

PhD THESIS

ISSN 1213-2365

Improving perception and locomotion capabilities of mobile robots in urban search and rescue missions

Vladimír Kubelka

kubelka.vladimir@cvut.cz

CTU–CMP–2018–01

June 4, 2018

Thesis Advisor: Tomáš Svoboda, Michal Reinštein

The work was supported by the EC project EU-FP7-ICT-609763, by GAČR No. 14-13876S and by SGS15/081/OHK3/1T/13, SGS16/161/OHK3/2T/13 and SGS18/138/OHK3/2T/13 of the Czech Technical University in Prague.

Research Reports of CMP, Czech Technical University in Prague, No. 1, 2018

Published by

Center for Machine Perception, Department of Cybernetics
Faculty of Electrical Engineering, Czech Technical University
Technická 2, 166 27 Prague 6, Czech Republic
fax +420 2 2435 7385, phone +420 2 2435 7637, www: http://cmp.felk.cvut.cz

Improving perception and locomotion capabilities of mobile robots in urban search and rescue missions

Vladimír Kubelka

June 4, 2018

Ph.D. Programme:
Electrical Engineering and Information Technology

Branch of study:
Artificial Intelligence and Biocybernetics

Supervisor: doc. Ing. Tomáš Svoboda, Ph.D.
Supervisor-specialist: Ing. Michal Reinstein, Ph.D.

I would like to thank my supervisors Tomáš Svoboda and Michal Reinstein for guiding me through my doctoral studies. They helped me to focus on relevant problems, discussed approaches to solve them, and shared their insight into publishing achieved results. Big thanks goes to prof. Václav Hlaváč who gave me a chance to take part in the NIFTi and TRADR projects, which were a great source of exciting research topics.

This work would not have been possible without collaboration with my colleagues from the NIFTi and TRADR projects and from the Center for Machine Perception, especially Martin Pecka, Tomáš Petříček, Karel Zimmermann, François Pomerleau, Francis Colas and Lorenz Oswald.

Special thanks to Jakub Šimánek, my colleague from the Department of Measurement, with whom we carried out hundreds of experiments, implemented various localization algorithms and wrote two papers.

Finally, my parents, my sister Katka and dear Janča were a great support, thank you for that and for your patience with me.

Abstract

Deployment of mobile robots in search and rescue missions is a way to make job of human rescuers safer and more efficient. Such missions, however, require robots to be resilient to harsh conditions of natural disasters or human-inflicted accidents. They have to operate on unstable rough terrain, in confined spaces or in sensory-deprived environments filled with smoke or dust. Localization, a common task in mobile robotics which involves determining position and orientation with respect to a given coordinate frame, faces these conditions as well. In this thesis, we describe development of a localization system for tracked mobile robot intended for search and rescue missions. We present a proprioceptive 6-degrees-of-freedom localization system, which arose from the experimental comparison of several possible sensor fusion architectures. The system was modified to incorporate exteroceptive velocity measurements, which significantly improve accuracy by reducing a localization drift. A special attention was given to potential sensor outages and failures, to track slippage that inevitably occurs with this type of robots, to computational demands of the system and to different sampling rates sensory data arrive with. Additionally, we addressed the problem of kinematic models for tracked odometry on rough terrains containing vertical obstacles. Thanks to research projects the robot was designed for, we had access to training facilities used by fire brigades of Italy, Germany and Netherlands. Accuracy and robustness of proposed localization systems was tested in conditions closely resembling those seen in earthquake aftermath and industrial accidents. Datasets used to test our algorithms are publicly available and they are one of the contributions of this thesis. We form this thesis as a compilation of three published papers and one paper in review process.

Keywords:
mobile robot, search and rescue, localization, data fusion, Kalman filter, odometry.

Abstrakt

Nasazení mobilních robotů během zásahů záchranných složek je způsob, jak učinit práci záchranářů bezpečnější a efektivnější. Na roboty jsou ale při takovém použití kladeny vyšší nároky kvůli podmínkám, které při těchto událostech panují. Roboty se musejí pohybovat po nestabilních površích, ve stísněných prostorech nebo v kouři a prachu, což ztěžuje použití některých senzorů. Lokalizace, v robotice běžná úloha spočívající v určení polohy robotu vůči danému souřadnému systému, musí spolehlivě fungovat i za těchto ztížených podmínek. V této dizertační práci popisujeme vývoj lokalizačního systému pásového mobilního robotu, který je určen pro nasazení v případě zemětřesení nebo průmyslové havárie. Nejprve je předveden lokalizační systém, který vychází pouze z měření proprioceptivních senzorů a který vyvstal jako nejlepší varianta při porovnání několika možných uspořádání takového systému. Lokalizace je poté zpřesněna přidáním měření exteroceptivních senzorů, které zpomalují kumulaci nejistoty určení polohy robotu. Zvláštní pozornost je věnována možným výpadkům jednotlivých senzorických modalit, prokluzům pásů, které u tohoto typu robotů nevyhnutelně nastávají, výpočetním nárokům lokalizačního systému a rozdílným vzorkovacím frekvencím jednotlivých senzorů. Dále se věnujeme problému kinematických modelů pro přejíždění vertikálních překážek, což je další zdroj nepřesnosti při lokalizaci pásového robotu. Díky účasti na výzkumných projektech, jejichž členy byly hasičské sbory Itálie, Německa a Nizozemska, jsme měli přístup na cvičiště určená pro přípravu na zásahy během zemětřesení, průmyslových a dopravních nehod. Přesnost našeho lokalizačního systému jsme tedy testovali v podmínkách, které věrně napodobují ty skutečné. Soubory senzorických měření a referenčních poloh, které jsme vytvořili pro testování přesnosti lokalizace, jsou veřejně dostupné a považujeme je za jeden z přínosů naší práce. Tato dizertační práce má podobu souboru tří časopiseckých publikací a jednoho článku, který je v době jejího podání v recenzním řízení.

Klíčová slova:
mobilní robot, záchranné práce, pátrací práce, lokalizace, fúze dat, Kalmanův filtr, odometrie.

Contents

1. Introduction

This work focuses on a problem of *mobile robot localization* in harsh conditions of *search and rescue missions*. The term localization denotes determining the position of some object with respect to a defined coordinate frame [22]. The mobile robot is a robot capable of autonomous locomotion. The search and rescue missions are the core objective of *rescue robotics* which is devoted to enabling responders and other stakeholders to sense and act at a distance from the site of a disaster or extreme incident [19].

The search and rescue robot is expected to serve as a sensory and manipulation platform providing a situation awareness in areas inaccessible or dangerous for human rescuers (Fig. 1.1 shows an example of such environment). These robots have been mostly teleoperated so far. End users have sometimes explicitly required removing any autonomous functions [19, 20]. However, teleoperation demands a trained operator dedicated for this job. An effort has been made to develop, test and deploy mobile robots with some level of *shared autonomy* (e.g. projects *NIFTi*[1] and *TRADR*[2] we have been working on). The *shared autonomy* means that human operator controls actions of a robot (e.g. decides where to go). Still, the robot has basic autonomous abilities. The autonomy is ideally implemented in a way the operator accepts it as a natural part of the system. This reduces cognitive load related to robot control – for example, an automatic collision avoidance allows the operator to do his job instead of focusing on obstacles in the robot way.

The *localization* of mobile robots is a necessary prerequisite for many of these basic automatic abilities. The autonomous navigation uses a known position of the robot to steer it along a planned path, detectors of dangerous chemical substances insert detections into a map based on position estimates. The localization can also help the robot find its way back from the disaster site when a wireless or tether connection to the operator is lost. Even without any autonomy, knowing position of the robot with respect to a map leads to a better situation awareness.

[1] European Union FP7 project NIFTi (No. 247870; *http://www.nifti.eu*)
[2] European Union FP7 project TRADR (No. 609763; *http://www.tradr-project.eu*)

Figure 1.1. NIFTi ground robot exploring collapsed church in Mirandola, Italy, 2012. Wooden and stone debris, which filled its nave, are remains of its roof and ceiling.

Contrary to laboratory conditions, design of a rescue robot has to follow constraints posed by harsh environment [19]:

1. *Extreme terrains and operating conditions affect size, sensor performance, and pose general robot survivability constraints.*
2. *Ability to function in GPS- and wireless-denied environments.*
3. *Provision of appropriate human-robot interaction for operators and for victims.*

Design of the localization system is constrained mostly by points 1. and 2. On flat surfaces, three degrees of freedom are often sufficient for robot localization. They express a 2D position together with a heading. Contrary to that, a difficult terrain with various obstacles implies estimating position in three dimensions together with robot attitude (six degrees of freedom in total). See Fig. 1.2 which depicts NIFTi ground robot navigating through earthquake aftermath training site used by Italian fire brigade. Such environment contains many obstacles the robot has to pass. Moreover, presence of dust, smoke or fog affects performance of various sensors, making them less reliable or even unusable. External sources of position such as GPS are usually blocked by structures present in rescue scenarios (e.g. collapsed buildings as earthquake aftermath or steel structures in industrial accidents).

We approach the problem of designing localization systems for harsh environments by investigating various *proprioceptive* and *exteroceptive*[3] sensory modalities. Proprioceptive modalities are more reliable especially in low-visibility conditions, exteroceptive modalities usually provide measurements which are more *information-rich*. For example, it is possible to estimate motion of a robot solely from a series of subsequent camera images (including translation scale given size of some observed objects is known). Sensors of an angular rotation rate, wheel motion encoders and several other sensors are needed to obtain similar results relying only on proprioceptive modalities. Cameras, however, struggle in areas filled with thick smoke.

Information provided by sensory modalities needs to be processed and fused to obtain estimates of a position and an attitude of the robot. We explore and propose fusion schemes that do this efficiently, preferably exploiting complementary character[4]

[3] Proprioception stands for modalities internal to the robot, e.g. inertial measurements, joint sensors, motor or wheel encoders, etc. Contrary to those, exteroception involves measuring properties external to the robot by cameras, laser range finders, sonars, etc.

[4] For example, estimating rotation rate from track velocities is not reliable but can be corrected by using an angular rate sensor instead.

Figure 1.2. NIFTi ground robot at a training site of the Italian fire brigade. An old military hospital has been partially torn down to prepare firemen for search and rescue operations.

of the modalities. To make localization more robust in real-world conditions, we consider methods of machine learning which, based on measured data, improve models of robot interaction with the environment and provide better interpretation of sensor measurements.

Important aspect of our work is identifying weak points of designed localization systems by testing them in conditions, which are likely to impair their functionality. After initial experiments in controlled laboratory environment, we proceed to terrains and locations resembling those observed in real search and rescue missions. They allow us to test accuracy, precision and robustness of localization in a way which is relevant to purpose of a search and rescue robot. Among these locations, the most challenging are training sites run by fire brigades. The ones we had access to are intended for training rescuers for earthquake and industrial accident scenarios (as shown in Fig. 1.4).

This thesis is a compilation of three published journal papers and one submitted journal paper in review. They are supported by work not included in this thesis but closely related; the work is referenced in Section 2. The three published papers are referenced as EKF Architectures [A.1], Data Fusion [A.2] and EKF Smoothing [A.4] further in the text to help the reader distinguish between them. The work included in this thesis proposes solutions to problems arising from development of a localization system for a tracked search and rescue robot (Fig. 1.3). This robot was developed as a part of EU funded projects *NIFTi* and *TRADR*. Their scope has been deployment of ground and air robots in search and rescue missions.

The localization architecture presented in this thesis starts as a proprioceptive data fusion scheme, which arose from need for a system estimating position and attitude of the tracked robot. Computational demands had to be limited allowing it to run on-board in a real time. We have decided to experimentally compare several state-of-the-art Kalman-filter-based architectures. Implementation of the best-performing

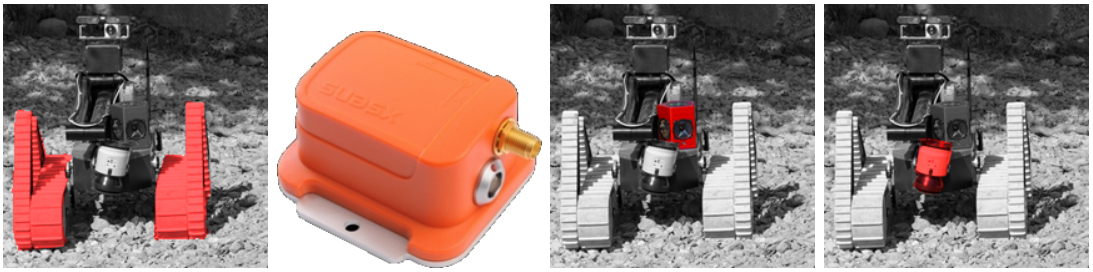


Figure 1.3. Sensory modalities available on our mobile robot for localization purposes. From the left to the right: Track odometry, inertial measurement unit, omni-directional camera and laser range-finder.

Figure 1.4. Blast furnace Phoenix West, Dortmund, Germany. Iron production has ceased and the site is used by local fire brigade to train for industrial accident scenarios. The TRADR project used this compound for end-user evaluations of the system.

fusion scheme has been deployed on the robot and further thoroughly tested during NIFTi and TRADR project end-user evaluations in real-world conditions.

Thanks to development of advanced exteroceptive sensory modalities provided by the project partners, the localization architecture has been extended to incorporate these additional measurements. Proposed EKF fusion scheme has to deal with varying sampling rates of the new exteroceptive modalities. Emphasis is put on the slowest one, which estimates relative motion of the robot by comparing consecutive 3D laser scans arriving every three seconds. We show how to apply corrections arising from these slow but accurate measurements while preserving trajectory details captured by higher-rate modalities.

Finally, for environments with difficult terrain, we have proposed additional kinematic models for tracked mobile robots, which improve localization accuracy during obstacle traversal. Complying to expected conditions of search and rescue missions, this improved odometry relies on proprioceptive data only making it usable in low-visibility environments. We have achieved reduction of localization drift, which otherwise deteriorates higher-level functionalities such as simultaneous localization and mapping.

2. Contributions

In this section, we present main contributions of the four journal articles that constitute the thesis core. We outline tasks we solved in the papers and provide a short summary of the results.

2.1. Extended Kalman Filter architectures for tracked mobile robots

The Extended Kalman Filter is a well-proven framework for data fusion and state estimation. However, there are several possible EKF architectures to choose from when designing a proprioceptive localization system for a tracked robot. When designing this system for the tracked NIFTi robot, we were unable to find any publication which would compare EKF architectures suitable for proprioceptive localization utilizing IMU and track odometry.

In our publication EKF Architectures [A.1], we decided to implement and experimentally compare four different state estimation architectures based on the EKF for 6-degrees-of-freedom proprioceptive localization. These architectures differ in definition of a system state, which is estimated by the EKF, and in the way the robot attitude is handled. The possible choices of the state reflected state-of-the-art approaches to inertial navigation and track odometry. The attitude estimation was extension of our previous results from [A.5]. Based on the achieved localization accuracy and computational demands of the tested architectures, we have found the best candidate for the task (see Fig. 2.1). The accuracy is approximately 4% position error of the distance driven, which is enough for local robot control and for higher-level exteroceptive localization modules to provide continuous corrections.

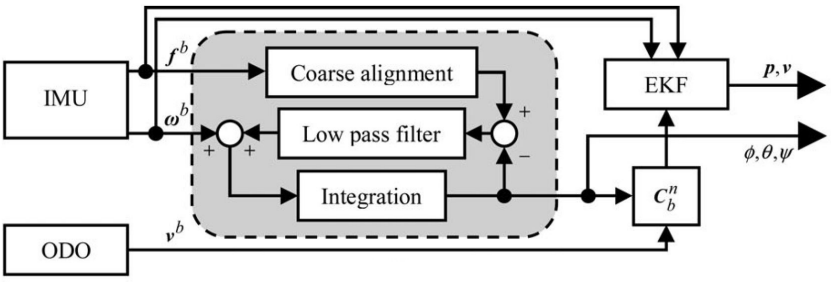


Figure 2.1. Best-performing EKF architecture selected from four alternatives. Data from inertial measurement unit (IMU) and track odometry (ODO) are fused to estimate the position, velocity and attitude expressed by three Euler angles. Adopted from EKF Architectures [A.1].

2.2. Multimodal data fusion for mobile tracked robots

Advances in sensor and computer technology allows mobile robots to be equipped with rich sensor suites and to process all measurements onboard. For purpose of localization, it is desirable to fuse all sensor modalities that capture the robot motion in a way that exploits their complementarity and redundancy. However, due to different capabilities of each sensor, measurements arrive at different sampling rates. Moreover, noise which they contain cannot be always simply modeled by a Gaussian distribution. The desired fusion system should take these characteristics into account, add resiliency to modality failures if redundancy allows that and provide robust localization estimate.

Fusion system proposed in Data Fusion [A.2] is based on the EKF and is our first contribution. It can fuse measurements with sampling rates spanning range from 0.3 to 100 Hz (see Fig. 2.2). A special attention is paid to 3D laser scan matching, which is accurate but difficult to integrate due to its extremely slow measurement rate. Three possible methods to achieve that were tested, one of them being our new proposed one. Experiments showed that our approach preserved details provided by faster sampling modalities while benefiting from corrections from the slow laser scan matching by reducing localization drift. The second contribution is the indoor and outdoor public dataset[1] available to the robotic community. It allows testing visual-based and laser-based localization techniques thanks to accurate reference recorded together with all sensor data. The dataset contains special experiments which invoke sensor failures intentionally and push the sensor modalities to their limits. The third publication EKF Smoothing [A.4] extends findings of Data Fusion [A.2] and compares proposed trajectory approach to another alternative technique, to the Rauch-Tung-Striebel smoother. Results show that our trajectory approach allows removal of drift caused by incorrect attitude present in laser-based localization modality.

2.3. Proprioceptive localization for tracked robots on vertical obstacles

Approaches to *track odometry* focus mainly on the slippage problem while turning. The task of traversing vertical obstacles has not been addressed in the context of localization. This poses a problem for robots operating in difficult terrains, especially in conditions that include smoke and dust, which complicates usage of additional exteroceptive sources of localization.

We proposed additional kinematic models for tracked vehicles traversing vertical obstacles, one of them is shown in Fig. 2.3. To switch between them, we have adopted a data-driven approach based on proprioceptive features thus making the odometry usable in low-visibility conditions. We have tested the proposed odometry system on artificial obstacles in a lab equipped with a tracking system. Recorded reference has allowed us to train the classifier for model switching. After initial experiments, we have subsequently proceeded to real-world obstacles, which the robot encounters in search and rescue scenarios. Applying the additional models, drift in z axis has been reduced. This is eventually beneficial for higher-level functionalities (e.g. laser mapping, path planning and execution), which otherwise suffer from gradual mis-localization (e.g. maps contain non-existent artifacts). We have submitted our findings as a journal paper, which is included as Section 6.

[1] The dataset is available at *https://sites.google.com/site/kubelvla/public-datasets*

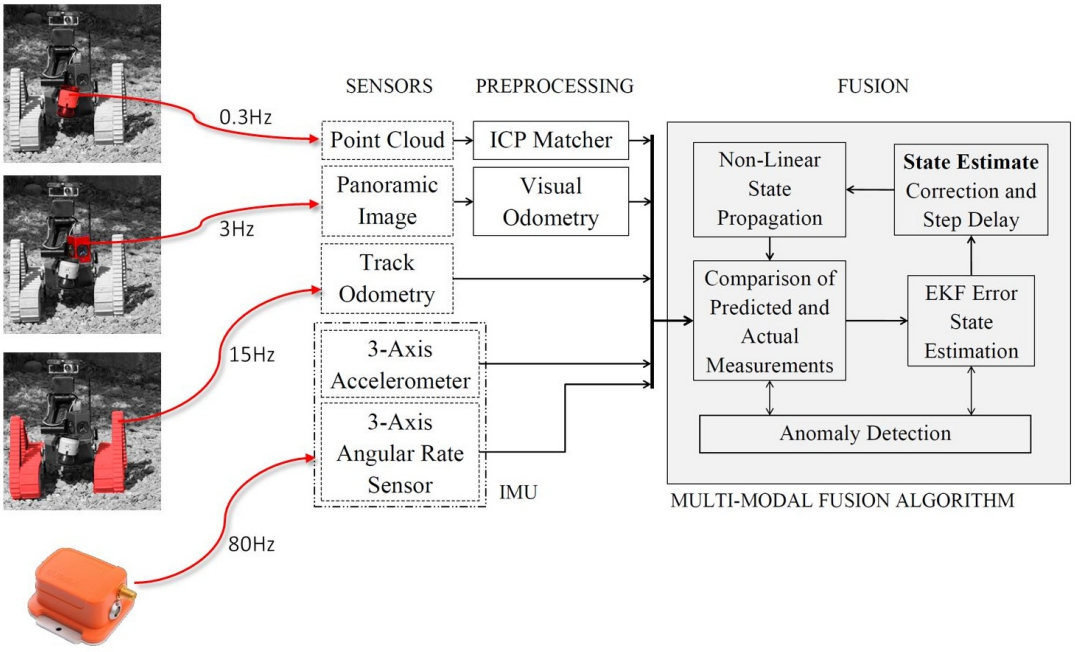


Figure 2.2. Schematics of the proposed data fusion for robot localization. Exteroceptive sensor data are preprocessed into measurements of position increments and fused together with the odometry and IMU data by the Extended Kalman Filter. The anomaly detection block is the result of our work [A.3].

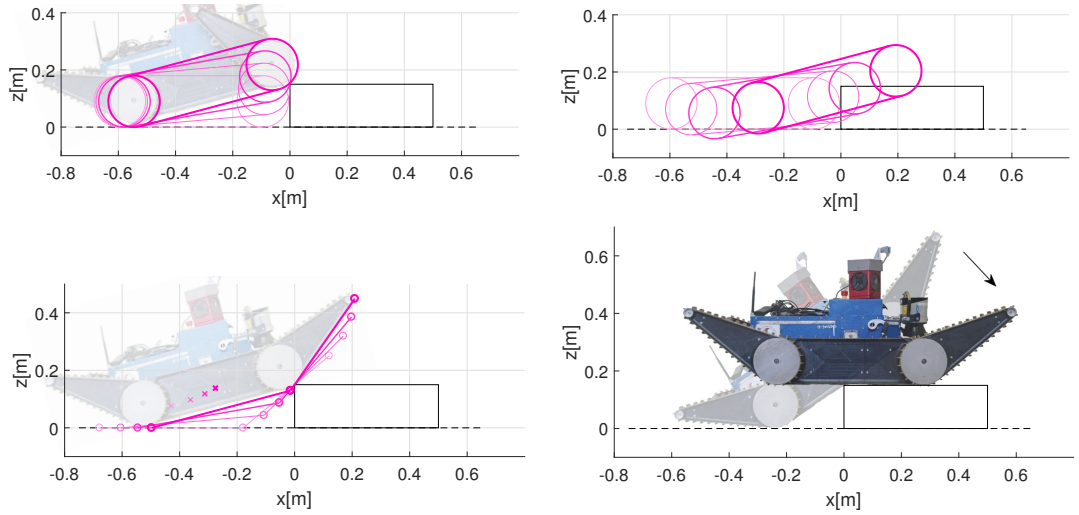


Figure 2.3. Kinematic model of a tracked robot climbing up a vertical obstacle. Top left: actual motion when climbing up an obstacle. Top right: trajectory estimated by a standard 3D odometry. Bottom left: proposed model for obstacle climbing with an adjustable sub-track. Bottom right: model of motion when tipping over the obstacle edge.

2.4. Related work

2.4.1. Related publications

Each paper in Sections 3, 4, 5 and 6 discusses in detail its state-of-the-art publications and our previous work. Additional references are provided in Section 7 which analyses topics related to search and rescue scenarios and advances in the field after our publications. Here, we provide an overview of topics discussed in Sections 3, 4, 5 and 6.

The first publication EKF Architectures [A.1] in Section 3 discusses work related to Extended Kalman filter (EKF) and architectures which utilize EKF to estimate position and attitude of a robot or of an aircraft. Publications from area of inertial navigation and attitude estimation are related closely. Their results provide mathematical model describing the relation between sensory measurements and the state of the system in a form of differential equations. Another modality of the dead reckoning[2] setup considered in the first publication is the track odometry, i.e. estimating the robot motion from its track velocities. We cite works which focus on the track slippage problem and on models for the slippage correction.

Our second publication Data Fusion [A.2] in Section 4 provides an overview of work related to four sensory modalities, which are available on our mobile robot platform and which provide localization-related measurements. Those are inertial measurements, track odometry, visual odometry and laser scan matching system. Fundamental works related to inertial navigation and dead reckoning partially overlap with those works cited in EKF Architectures [A.1]. Additional publications give examples of processing relative motion measurements. We further introduce publications from the field of visual odometry and laser scan matching, localization and mapping.

The third publication EKF Smoothing [A.4] in Section 5 extends findings of Data Fusion [A.2]. Therefore its related work section recaps fundamental works from Data Fusion [A.2] and adds publications related to Kalman filter smoothers and to smoothing techniques, which involve the graph optimization and inverse filters for visual-inertial systems.

Section with related work in the last article, included in Section 6, which has been submitted for review, provides an overview of the problem of estimating motion of a tracked robot while traversing obstacles. The section shows the approaches first which involve mechanical modifications of the robot in order to improve ability to climb and traverse the obstacles. These modifications include adjustable morphology of the robot. After discussing mechanical approaches, works providing means to analyze the ability to traverse a given obstacle are introduced. Publications proposing ways to correctly estimate motion of the robot based on track velocities are the most related to our work. Additionally, we cite several publications which propose various sensors that improve perception of the terrain beneath the tracks.

2.4.2. Related projects

Two search and rescue projects NIFTi and TRADR motivated us to investigate localization for mobile robots. Their main objective was formation of mixed human-robot rescue teams. Real-world rescue scenarios context and a strong emphasis on the user-centric design have affected development of software and hardware within these projects.

[2] Term *dead reckoning* is used in navigation for determining one's current position based on previously estimated position. The new position estimate is obtained by integrating speed and course measurements.

Proposed algorithms have been tested in the same environments which human rescuer teams use for training. They define limitations needed to be taken into account:

- The localization is required to run on-board the mobile robot because of limited wireless connection to the operator side.
- Sensor measurements are noisy. Exteroceptive sensor measurements can be obfuscated by smoke and dust.
- Proprioceptive sensors such as inertial sensors suffer from vibrations and strong jolts when robot moves on a rough terrain.

Software design has benefited from using Robot Operating System simplifying integration of different sensory modalities, fusion systems and higher-level functions. In the case of the localization system, we have aimed for modular, loosely coupled systems. This has allowed us to upgrade the system components without major code changes to the rest of the modules in the eight years of the two projects time span. This choice comes with a potential loss of optimality, of course.

Proposed algorithms have been also tested in two real deployments. Both took place in Italy in the earthquake aftermath. The first mission was in Mirandola in year 2012, where NIFTi robots helped to map state of a damaged church, which was marked as a no-go zone for human rescuers. Video and depth data were captured helping with damage and danger assessment. Due to early stage of the NIFTi project with software and hardware not thoroughly tested, the robot was tele-operated by a human operator maintaining visual contact through entrance to the church. The second real-world deployment took place in Amatrice in 2016. The town was struck by another earthquake, two ground robots and several UAVs were deployed to fulfill a mission similar to Mirandola – damage assessment and creation of textured 3D models of two churches including their interior (Fig. 2.4, 2.5).

The ground robots were running implementations of the localization discussed in EKF Architectures [A.1] and Data Fusion [A.2] helping mainly in the second Amatrice mission where visual contact with the robots was not possible. The operators had to rely on 3D visualizations of robot surroundings in a form of a 3D *point clouds*[3], which were continuously assembled based on known position and attitude of the robot. Video feed was available as well but found insufficient for assessing size of obstacles in robot way.

[3]The point cloud is a representation of a set of laser range-finder or another depth-measuring sensor measurements. It is simply a set of 3D points.

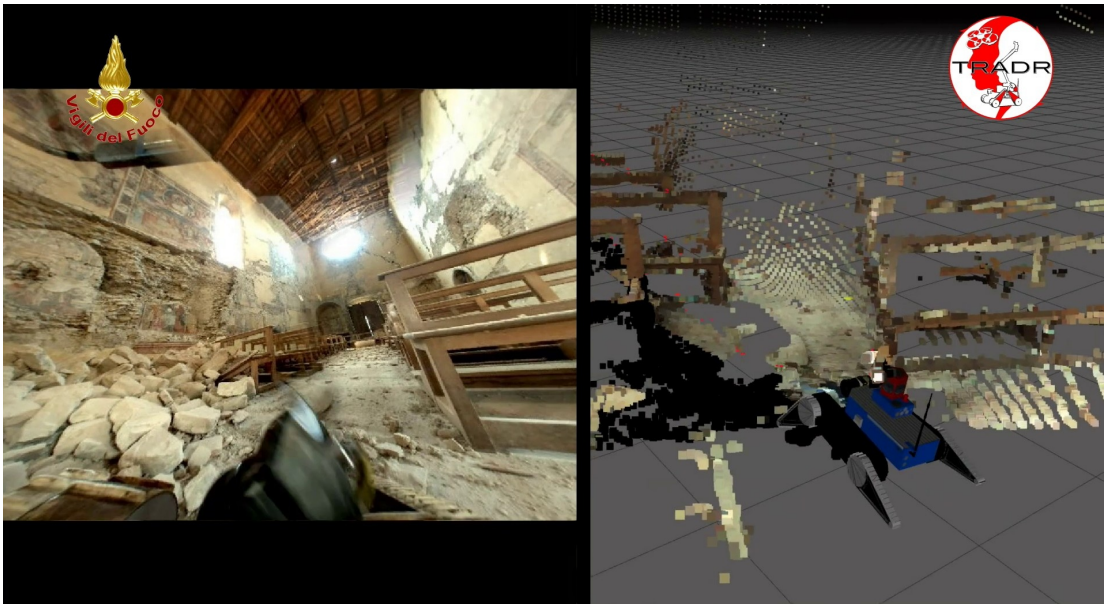


Figure 2.4. A view from the on-board camera (left). A 3D colored point cloud (right), both captured in a collapsed church in Amatrice, Italy.

Figure 2.5. Italian firefighters assisting with the robot retrieval from a no-go zone after returning from the collapsed church. Amatrice, Italy.

3. Evaluation of the EKF-based estimation architectures for data fusion in mobile robots

This section presents the first publication, the EKF Architectures [A.1]. It is a 6-page-long paper published in *IEEE/ASME Transactions on Mechatronics*. Here, we give a short overview of the paper and then provide the paper itself as the last part of this chapter.

The **introduction** section briefly establishes context of search and rescue robotics and then moves to description of Extended Kalman Filter architectures which are possible candidates for a dead-reckoning localization system. It introduces linear and non-linear kinematic models with references to related papers and complementary filter for attitude estimation [A.5]. Resulting four architectures are examined in the remainder of the paper.

The **theory** section presents standard EKF equations for discrete-time systems and then puts it in context of the two examined system models and two ways of estimating robot attitude. The measurement model based on the track odometry is defined and observability analysis is performed, giving the reader idea about uncertainty of the state estimate and its expected development in time. Moreover, consistency tests are performed for the discussed architectures, checking whether corrections applied to state estimate conform to expected statistical distribution.

The next section describes **experiments** performed in order to compare the accuracy and computational demands of the four competing EKF dead-reckoning localization architectures.

After more **implementation details** are given regarding the best-performing architecture running on-board the robot under the Robot Operating System[1] (ROS), our findings are **concluded** in the last section.

[1] Robot Operating System, *http://www.ros.org*

Evaluation of the EKF-Based Estimation Architectures for Data Fusion in Mobile Robots

Jakub Simanek, Michal Reinstein, and Vladimir Kubelka

***Abstract*—This paper presents evaluation of four different state estimation architectures exploiting the extended Kalman filter (EKF) for 6-DOF dead reckoning of a mobile robot. The EKF is a well proven and commonly used technique for fusion of inertial data and robot's odometry. However, different approaches to designing the architecture of the state estimator lead to different performance and computational demands. While seeking the best possible solution for the mobile robot, the nonlinear model and the error model are addressed, both with and without a complementary filter for attitude estimation. The performance is determined experimentally by means of precision of both indoor and outdoor navigation, including complex-structured environment such as stairs and rough terrain. According to the evaluation, the nonlinear model combined with the complementary filter is selected as a best candidate (reaching 0.8 m RMSE and average of 4% return position error (RPE) of distance driven) and implemented for real-time onboard processing during a rescue mission deployment.**

***Index Terms*—Complementary filter (CF), extended Kalman filter (EKF), urban search and rescue (USAR).**

I. Introduction

The recent progress in sensor technologies and increase in onboard computational power brings new demands and pushes the limits of navigation of autonomous robots. Mobile robots are becoming increasingly more reliable and hence more popular even for complex missions such as urban search and rescue (USAR). In general, there exist a number of different solutions to data fusion for localization and navigation, making it often unclear when deciding for one that ensures optimal performance. Moreover, the performance is influenced not only by the choice of the platform, its morphology, and the sensor suite, but also by environment and mission specifications. Development of the system for USAR missions aims primarily on human–robot teaming, especially deploying an unmanned ground vehicle (UGV) (see skid-steer robot UGV in Fig. 1) in a close cooperation with professional rescue teams. The mission, in which the testing UGV was deployed and used one of the EKF architectures proposed in this paper, consisted of reconnaissance of earthquake-affected areas[1] (refer to [1] and [2] for details about the USAR challenge and mission, respectively). In order to ensure reliable performance, explore the disaster site, and perform high-quality 3-D mapping of the environment (such as in [3]) reliable dead reckoning development was essential.

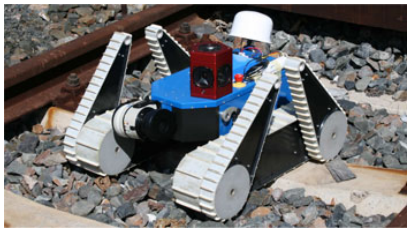


Fig. 1. Skid-steer mobile robot (the UGV) for search and rescue operations developed by BlueBotics (www.bluebotics.com) as part of NIFTi project.

Since the extended Kalman filter (EKF) is a well proven and commonly used state estimation technique [4], [5], the motivation for this paper lied mainly in seeking the most suitable EKF architecture. To our best knowledge, we are convinced no such comparison of estimation architectures covering EKF-based dead reckoning was carried out for mobile skid-steer robots intended for real USAR missions. Thus, the results of the analysis might prove beneficial to anyone asking the fundamental question: What is actually the best way to implement the EKF? Therefore, objective of this paper lies in comparing four different approaches: a nonlinear model (NLM) [6] and an error model (ERM) [7], each with and without a complementary filter (CF) for attitude estimation [8], [9]. The performance of attitude estimation using the CF was thoroughly evaluated as part of our previous work in [8], including testing of various filters to cope with inertial signals strongly affected by vibration. In this paper, dead reckoning is represented by estimation of the six-degree-of-freedom (6-DOF) pose of the UGV, using the proprioceptive sensors only: odometry obtained from motor encoders and an inertial measurement unit (IMU) that consists of accelerometers and gyroscopes providing specific force and angular rate measurements [10]. As it was shown repeatedly, the combination of IMU and odometry is a popular technique to localize a mobile robot even in case of dynamic- legged robots [11]. It generally allows high sampling and processing rate, usually without excessive computational load. However, there are drawbacks, such as inertial navigation drift and wheel slip [12], which also reflect to the principle of dead reckoning. Therefore, substantial effort has been made to investigate improvements through slip estimation, velocity constraints [13], [14], odometry-derived constraints [15], or innovative motion models [16]. We have also addressed the slip compensation problem in [17]; however, our aim is not to develop a dead reckoning with bounded position error, which is in principle not achievable using proprioceptive sensors only [18]. Therefore, most of

Manuscript received November 2, 2012; revised February 16, 2013 and October 23, 2013; accepted February 27, 2014. Date of publication April 1, 2014; date of current version October 24, 2014. Recommended by Technical Editor Y. Li. This work was supported by the EC project FP7-ICT-247870 NIFTi and work on the last revision was supported by Project TRADR FP7-ICT-609763.

J. Simanek is with the Department of Measurement, Faculty of Electrical Engineering, Czech Technical University in Prague, 166 36 Prague, Czech Republic (e-mail: simanjak@fel.cvut.cz).

M. Reinstein and V. Kubelka are with the Center for Machine Perception, Department of Cybernetics, Faculty of Electrical Engineering, Czech Technical University in Prague, 166 36 Prague, Czech Republic (e-mail: reinstein.michal@fel.cvut.cz; kubelvla@fel.cvut.cz).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TMECH.2014.2311416

[1]For more details see the European Commission Memo http://europa.eu/rapid/press-release_MEMO-12-620-en.htm

TABLE I
NOMENCLATURE

$b; x, y, z$	body-frame; origin: UGV's center of mass; x-axis pointing forwards, y-axis pointing right, z-axis pointing down
$n; N, E, D$	local navigation-frame; origin: UGV's initial position; N-axis pointing north, E-axis pointing east, D-axis pointing down
$\boldsymbol{x}, \boldsymbol{P}$	state vector, estimation error covariance matrix
$x^+, x^-, \delta x$	*a priori* value; *a posteriori* value; error value of x
$\boldsymbol{u}, \boldsymbol{z}$	input and measurement vector
f, h	non-linear process and measurement model function
$\boldsymbol{F}, \boldsymbol{H}; \boldsymbol{K}$	partial derivative matrices (Jacobians); Kalman gain matrix
$\boldsymbol{w}, \boldsymbol{v}; \boldsymbol{Q}, \boldsymbol{R}$	process and measurement: noise vectors; covariance matrices
$k; \boldsymbol{0}_{3\times3}, \boldsymbol{I}_{3\times3}$	discrete time step; 3×3 zero and identity matrix
$\boldsymbol{p}, \boldsymbol{v}, \boldsymbol{a}$	position, velocity and attitude vector (m; m/s; deg/s)
$\boldsymbol{f}^b, \boldsymbol{\omega}^b$	specific force and angular rate vector (m/s^2; deg/s)
ϕ, θ, ψ	Euler angles: roll, pitch, yaw (deg)
$\boldsymbol{C}_b^n$	body to navigation frame transformation matrix

the state-of-the-art 6-DOF IMU and odometry dead reckoning solutions require other supporting sensor systems, especially exteroceptive, to assure desired precision and reliability given by the target application; for example, see [18]–[21].

The contributions of this paper are: first, providing the reader with analysis regarding the performance of different architectures for state estimation using the EKF; second, extending our previous work about the CF for attitude estimation [8] to full 6-DOF dead reckoning; third, testing in both indoor- and outdoor-structured environment (including stairs and other obstacles); and fourth, implementing the best approach for onboard processing in the Robot Operating System (ROS) [22].

This paper is structured as follows. Section II covers details regarding the theory and methodology. Section III presents experimental results and performance evaluation. Section IV provides implementation details. Conclusions are given in Section V. All the symbols used can be found in Table I.

II. Theory and Methodology

A. Extended Kalman Filter

The EKF is a linearized extension of the Kalman filter that estimates the states of a process and their uncertainty from noisy data. The EKF algorithm was implemented using the standard equations (see [23] for more details and relevant notation in Table I). Assuming a discrete-time system model disturbed by normally distributed noise

$$\boldsymbol{x}_k = f_{k-1}\left(\boldsymbol{x}_{k-1}, \boldsymbol{u}_{k-1}\right) + \boldsymbol{w}_{k-1}, \boldsymbol{w}_{k-1} \sim N\left(0, \boldsymbol{Q}_k\right) \quad (1)$$

$$\boldsymbol{z}_k = h_k\left(\boldsymbol{x}_k\right) + \boldsymbol{v}_k, \boldsymbol{v}_k \sim N\left(0, \boldsymbol{R}_k\right). \quad (2)$$

EKF is initialized with the following expected values of the estimated state and covariance:

$$\boldsymbol{x}_0^+ = E\left[\boldsymbol{x}_0^+\right], \boldsymbol{P}_0^+ = E\left[(\boldsymbol{x}_0 - \boldsymbol{x}_0^+)(\boldsymbol{x}_0 - \boldsymbol{x}_0^+)^{\mathrm{T}}\right]. \quad (3)$$

The EKF algorithm can be described in two steps using the following discrete-time equations.

1) *Time update:* Predicted (*a priori*) state is computed; state transition Jacobian is determined to evaluate the predicted covariance

$$\boldsymbol{x}_k^- = f_{k-1}\left(\boldsymbol{x}_{k-1}^+, \boldsymbol{u}_{k-1}\right), \boldsymbol{F}_{k-1} \approx \left.\frac{\partial f_k}{\partial \boldsymbol{x}}\right|_{\boldsymbol{x}_{k-1}^+, \boldsymbol{u}_{k-1}} \quad (4)$$

$$\boldsymbol{P}_k^- = \boldsymbol{F}_{k-1}\boldsymbol{P}_{k-1}^+\boldsymbol{F}_{k-1}^{\mathrm{T}} + \boldsymbol{Q}_{k-1}. \quad (5)$$

2) *Measurement update:* Observation Jacobian and Kalman filter gain are computed; *a priori* estimates with the measurements are combined to provide the *a posteriori* state and covariance estimates

$$\boldsymbol{H}_k \approx \left.\frac{\partial h_k}{\partial \boldsymbol{x}}\right|_{\boldsymbol{x}_k^-}, \boldsymbol{K}_k = \boldsymbol{P}_k^-\boldsymbol{H}_k^{\mathrm{T}}\left(\boldsymbol{H}_k\boldsymbol{P}_k^-\boldsymbol{H}_k^{\mathrm{T}} + \boldsymbol{R}_k\right)^{-1} \quad (6)$$

$$\boldsymbol{x}_k^+ = \boldsymbol{x}_k^- + \boldsymbol{K}_k\left(\boldsymbol{z}_k - h_k\left(\boldsymbol{x}_k^-\right)\right), \boldsymbol{P}_k^+ = \left(\boldsymbol{I} - \boldsymbol{K}_k\boldsymbol{H}_k\right)\boldsymbol{P}_k^-. \quad (7)$$

The EKF relies on assumption of an approximate equality of the estimated state between the two consecutive time steps. The state prediction and update propagates through the nonlinear system functions, and the state and observation errors propagate through a separate linearized system, which is formulated as Taylor series about the estimate.

B. Process Models

Four different schemes for data fusion using EKF are presented in this paper and shown in Fig. 2 (see relevant notation in Table I). Two different system process modeling approaches were investigated—the NLM and the ERM, see Fig. 2(a) and (c) for implementation without the CF (standard solutions), and Fig. 2(b) and (d) for the implementation with the CF for attitude estimation (denoted as the grey block in Fig. 2). The CF was introduced in [8] and proved to be an effective algorithm, where the attitude determination is based on fusion of the gravity vector determined from specific forces via a coarse alignment algorithm (see [10]), and Euler angles (roll, pitch, and yaw) computed as integration of angular rates.

First, the NLM (for details see [6]) is based on coordinate frames transformation (where the nonlinearities are caused by the direction cosine matrix) and numerical integration of the IMU outputs. It handles navigation states (position in navigation frame, velocity in body-frame, and Euler angles) directly. For the purpose of UGV navigation, the centripetal acceleration corrections can be neglected.

Second, the ERM, derived by linearizing the nonlinear differential equations using perturbation analysis and thoroughly described in [7], is based on a 15-state concept, expanding the estimates of position and velocity in the navigation frame, and attitude errors (i.e., the difference from the expected value) with the biases of the inertial sensors (the actual sensor errors). The data fusion must be completed with an error control loop, which provides corrections to the navigation states computed using the differential navigation equations named as *mechanization* by Rogers in [24]. Unlike in the case of the NLM, it is necessary to implement the mechanization, because the ERM provides only corrections, not the actual navigation states. Every time the corrections are fed back to the mechanization, the EKF state vector must be set to zeroes.

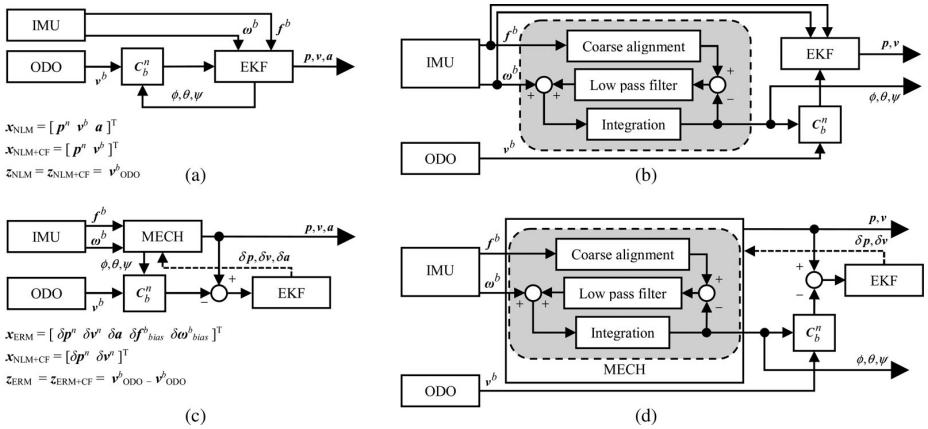


Fig. 2. IMU and odometry (ODO) data fusion schemes for the EKF: (a) with the NLM, (b) with the NLM and CF for attitude estimation (NLM+CF), (c) with mechanization (MECH) and the ERM, and (d) with mechanization (MECH), ERM, and CF for attitude estimation (ERM+CF); $\boldsymbol{x}$ and $\boldsymbol{z}$ are the state and measurement vectors. (a) NLM. (b) NLM + CF. (c) ERM. (d) ERM + CF.

Considering the solutions enhanced by the proposed CF [8], which provides stable estimates in pitch and roll channels, there is no need to estimate the attitude angles (or attitude corrections) and biases of the gyroscopes within the EKF model, since the feedback of the CF assures stability and drift suppression. Therefore, in both the nonlinear and the ERM with the CF, the state vector consists of position and velocity estimates or their respective error estimates.

C. Measurement Model

The measurement model is based on the standard body frame odometry model transformed to the navigation frame (see Table I for details about frames). Since the UGV is bounded to the surface, it is said to be governed by nonholonomic constraints, which can be exploited to aid the estimation of IMU alignment [15]. These constraints are incorporated into the measurement model, such that no side slip and no vertical movement (such as free fall) are assumed along the body frame lateral y-axis and vertical z-axis, respectively. Therefore, 3-D odometry model is implemented as follows: the tracks and UGV's body are aligned; therefore, the navigation frame velocity is computed as

$$\boldsymbol{v}_k^n = \boldsymbol{C}_b^n \boldsymbol{v}_k^b = \boldsymbol{C}_b^n \begin{bmatrix} 0.5\,(\boldsymbol{v}_L + \boldsymbol{v}_R) & 0 & 0 \end{bmatrix}^{\mathrm{T}} \tag{8}$$

where v_L and v_R are the velocities of the left and right tracks, respectively, and the nonholonomic constraints are incorporated as zero values in the y- and z-component of the displacement vector. The body to navigation frame transformation matrix is defined as

$$\boldsymbol{C}_b^n = \begin{bmatrix} c_\theta c_\psi & s_\phi s_\theta c_\psi - c_\phi s_\psi & c_\phi s_\theta c_\psi + s_\phi s_\psi \\ c_\theta s_\psi & s_\phi s_\theta s_\psi + c_\phi c_\psi & c_\phi s_\theta s_\psi - s_\phi c_\psi \\ -s_\theta & s_\phi c_\theta & c_\phi c_\theta \end{bmatrix} \tag{9}$$

where $\phi, \theta,$ and ψ are the roll, pitch, and yaw angles (Euler angles), respectively, and $c_\theta = \cos\theta$, $s_\theta = \sin\theta$, etc.

The EKF measurement vector for both the NLM and the NLM+CF [see Fig. 2(a)] consists of velocity in the navigation frame. In case of the ERM and the ERM+CF, the measurement vector is produced as the difference between the velocity computed by the mechanization, and odometry-determined velocity, both expressed in navigation frame, as shown in Fig. 2(c). However, the wheel slip along the body frame x-axis is still apparent, and the model cannot compensate for it.

D. Observability Analysis

Observability analysis determines, whether a state can be estimated from available measurements [25]. Observability rank tests were implemented according to [7] and [26] along with the inspection of the behavior of the estimation error covariance matrix. The observability analysis of the nine navigation states in both the NLM and the ERM revealed the following conclusions about the attitude (also corresponds to results in [26]): roll and pitch angles are observable with respect to the navigation frame due to gravity measurement provided by IMU; the yaw angle becomes unobservable with respect to the navigation frame when the system becomes stationary. Based on these conclusions, an odometry-derived motion constraint was introduced to reduce the drift in the unobservable yaw. As the odometry directly indicates when the UGV is stationary, an artificial limit to constant heading was utilized and proved to eliminate the drift in the unobservable yaw angle. Since the velocity is provided to the EKF as measurement, its covariance converges quickly; therefore, the velocity estimation error is bounded. This does not reflect the cases, when the UGV is exposed to maneuvers on harsh terrain causing slippage; however, compensating it is not aim of this study (we address it separately in [17]). Consequently, the position-related covariance diverges in all four approaches, as it should in any other dead reckoning approach that lacks (in principle) the absolute position measurements [18]. This is covered in detail in [23] and confirms that additional position measurements are necessary to achieve observable position estimates.

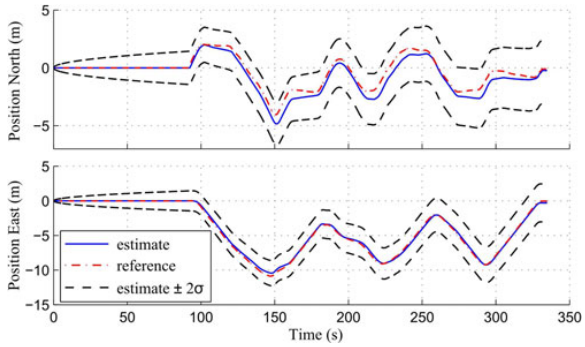


Fig. 3. Position and standard deviation σ estimates used for filter tuning with respect to the position reference; one of the 2-D tracked experiments as obtained using the NLM+CF approach.

E. Filter Initialization, Tuning, and Consistency

Initialization sequence requires the UGV to be stationary for some time. During this interval, the coarse alignment algorithm (see [10]) is used to determine roll and pitch angles from averaged specific forces. Along with the initial attitude, biases of the gyroscopes are estimated as the average of the static angular rate measurements. This requires at least 1 min averaging and optimally 3 min (providing appropriate bias estimates for 20 min experiments). Initial position of the UGV always corresponds with origin of the local navigation frame and initial velocity is set to zero, as well as their initial covariance matrices.

Filter consistency is defined as convergence of the estimate to the true value [27]. If the estimator is inconsistent, estimates are not reliable, because their accuracy is unknown. As the process of choosing noise covariance matrices (referred as *filter tuning*) is often based on the tradeoff between obtaining consistent filter and small estimation errors [27], the noise covariance matrix was tuned such that both real-time consistency tests involving the *zero-mean innovation test* and the *innovation whiteness test* [27] were passed. Subsequently, the process noise covariance matrix was adjusted to achieve small estimation error with respect to the ground truth. Consistency of the position estimate was checked via the inspection of the covariance and the progress of position error (see an example for a 2-D experiment using the NLM+CF approach in Fig. 3).

The summary of the EKF initialization parameters (state vector, error covariance matrix, process noise covariance matrix, and measurement noise covariance matrix) is given in Table V.

III. Experimental Evaluation

To evaluate the navigation performance of all four approaches, series of indoor and outdoor navigation experiments in complex-structured environment were performed. The testing UGV was a mobile skid-steer robot (see Fig. 1) designed for USAR operations and developed as a part of the NIFTi project concerning mainly human–robot cooperation. The UGV is equipped with the following sensors used in this approach: Xsens MTi-G unit (inertial data at 90 Hz) and motor encoders (left and right track velocities at 15 Hz). The evaluation of experiments was performed offline in MATLAB, and according to the results, the best performing architecture was then implemented in C++ for ROS.

TABLE II
Experiment Datasets

Experiment name	*2D Tracked*	*Hallway*	*Stairway*	*Outside*
Average distance (m)	63	110	90	195
Vertical difference (m)	0	0.3	5.0	2.0
Average duration (s)	180	450	750	1200

The experiments involved four types of field-testing environments—namely the *2-D tracked* outdoor experiments (planar; evaluated with respect to ground truth trajectories[2] available only for the *2-D tracked* dataset), indoor experiments on *hallway* (UGV exploring a ground floor corridor inside a building) and *stairway* (extended to climbing of wooden staircase and descending at slope of 20°), and *outside* environment (long-term driving through a natural environment including hillside and stair climbing). All datasets are referred to using the corresponding *labels*. In the case of 3-D experiments, ground truth was not available, and hence, the UGV was driven back to the exact starting position each time. Therefore, the return position error (RPE) and the return attitude error (RAE) were evaluated instead of RMSE with respect to a reference. The total distance travelled during the 2-D experiments was 2.2 km (35 experiments), and 2 km during the 3-D experiments (five experiments for each 3-D field-test). Average characteristics of each experiment type are concluded in Table II.

A. Example of Results: The Stairway Experiment

One of the typical datasets collected during the *stairway* experiment was chosen to demonstrate the performance by comparing the attitude and position estimates. In the *stairway* experiment, the UGV was driven forward from the initial location to the stairs, then the UGV climbed up to the second floor, turned left, continued forward to the turning point, and returned along a similar path to the initial position.

Attitude estimates that were obtained during the stair descent using all the approaches are shown in Fig. 4 (please note, that the NLM+CF and the ERM+CF exploit the same CF for attitude estimation, thus the figure shows only the CF attitude instead). There are minor differences in roll and pitch channels for the NLM and the CF approaches. Although the ERM performs locally inconsistently, overall stability is ensured and depends mainly on the filter tuning. The major contribution to the position error in dead reckoning rises from the yaw estimates. All approaches provide slightly different yaw angle during the stair descent and overall experiment. The consequence can be seen in the trajectory estimates shown in Fig. 5 (position projected to 2-D North-East and Down-East directions), where all the North-East horizontal turns should be nearly perpendicular. In the used

[2]The ground truth was obtained using top-down camera tracking of the UGV. The precision of the video tracking system was determined experimentally to be 15 cm ± 12 cm within a 15 m × 10 m outdoor area.

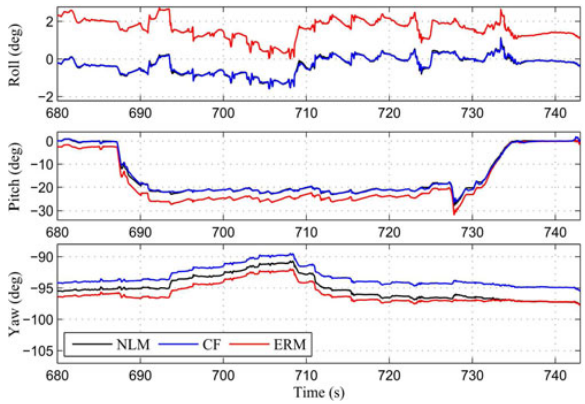


Fig. 4. Euler angles for the *stairway* experiment—stair descent (680 to 745 s zoom) as obtained using all attitude approaches.

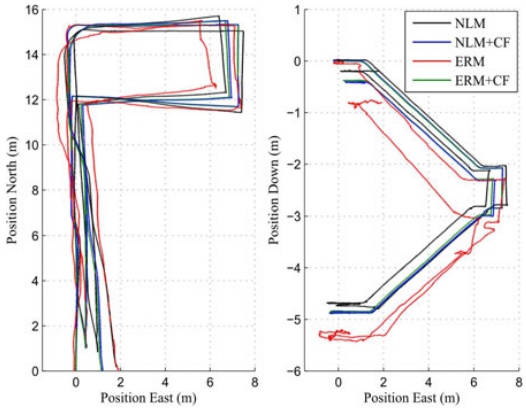


Fig. 5. North-East $(x-y)$ and Down-East $(z-y)$ projections of the 3-D trajectory for the *stairway* experiment as obtained using all four approaches.

dead reckoning, the overall trajectory precision depends mostly on attitude estimates. Another type of positioning error is apparent at the ground and first floor in the Down-East view in Fig. 5, where slipping of the tracks during climbing and descending also contributes to the different vertical position estimates.

B. Performance Evaluation

To demonstrate the relative performance, all the collected datasets were evaluated statistically, averaging, and concluding all the experiments in Table III. The RMSE in 2-D position (determined as the RMSE of the north and east position), the average RPE of the distance driven, and the average RAE of roll and pitch absolute differences, show that the approaches with the CF outperform the standard ones for both the NLM and the ERM implementation.

The computational load analysis of all the four approaches was also performed. The results in Table IV (evaluated in MATLAB 2012a running on PC with Intel Core i7 2.6 GHz, 8-GB RAM) indicate that the most demanding ERM with mechanization consumed approximately two times more computational time than the NLM. In the NLM+CF and the ERM+CF, the state vectors are reduced as discussed in Section II-B; thus, the EKF computational load is lower, especially for the ERM+CF. According to the results concluded in Tables III and IV, the NLM+CF approach was selected as the best candidate for the onboard processing in ROS Fuerte version [22].

TABLE III
AVERAGE SUMMARY OF THE PERFORMANCE EVALUATION: 2-D RMSE, RPE OF THE DISTANCE DRIVEN, RAE OF ROLL AND PITCH ANGLES

Approach	2D Tracked	Hallway		Stairway		Outside	
	RMSE (m)	RPE (%)	RAE (deg)	RPE (%)	RAE (deg)	RPE (%)	RAE (deg)
NLM	0.92	5.2	0.24	3.5	0.23	5.1	0.15
NLM+CF	0.79	4.1	0.21	3.3	0.23	4.7	0.13
ERM	0.98	4.8	0.70	3.1	0.58	5.8	0.65
ERM+CF	0.87	4.1	0.21	3.6	0.23	4.7	0.13

TABLE IV
COMPUTATIONAL LOAD ANALYSIS (MATLAB)

	NLM	NLM+CF	ERM	ERM+CF
EKF cycle duration (ms)	0.8	1.0	2.3	2.1

TABLE V
EKF INITIALIZATION PARAMETERS

	Nonlinear model	Error model
$\boldsymbol{x}_0$	$[\mathbf{0}_{6\times1} \;\; \phi_0 \;\; \theta_0 \;\; \psi_0]^T$	$[\mathbf{0}_{15\times1}]^T$
$\boldsymbol{P}_0$	$diag([\mathbf{0}_{6\times6} \;\; 0.05\times\boldsymbol{I}_{3\times3}])$	$diag([\mathbf{0}_{6\times6} \;\; 0.01\times\boldsymbol{I}_{9\times9}])$
$\boldsymbol{Q}_k$	$diag(10\times[\sigma^2_{acc}\times\boldsymbol{I}_{3\times3} \;\; \sigma^2_{gyr}\times\boldsymbol{I}_{3\times3}])$	$diag([\mathbf{0}_{9\times9} \;\; \sigma^2_{acc}\times\boldsymbol{I}_{3\times3} \;\; \sigma^2_{gyr}\times\boldsymbol{I}_{3\times3}])$
$\boldsymbol{R}_k$	$0.1\times\boldsymbol{I}_{3\times3}$	$10^{-3}\times\boldsymbol{I}_{3\times3}$

diag constructs block diagonal matrix; $\sigma^2_{acc} = 5\times10^{-4}$ m^2/s^4 and $\sigma^2_{gyr} = 5\times10^{-4}$ rad^{-2} stand for accelerometer and gyroscope noise variances determined experimentally from static measurements.

IV. IMPLEMENTATION DETAILS

Since the reader should benefit not only from the knowledge gained from presented experimental evaluation, this section provides more general remarks about the implementation.

The NLM can be considered as the most straightforward EKF navigation algorithm, based on suboptimal equations estimating position in the navigation frame, velocity in the bodyframe, and attitude in Euler angles. Thus, this process model ensures straightforward debugging and EKF tuning. The ERM and mechanization represents computationally demanding but precise navigation algorithm designed to cope with coning and sculling effects in inertial sensors, Coriolis and gravity corrections, Earth's rate, and transport rate (please consult terms with [7] and [10]). The reader should consider the mechanization with the ERM when desiring such complex solution providing position and velocity corrections in the navigation frame and attitude corrections as a direction cosine matrix compensation. On the other hand, the filter tuning has to be performed experimentally and is directly connected with the precision of the ground truth and number of experiments. Using inertial sensors only, the CF proved to provide stable and reliable roll and pitch angles for

the UGV platform, where no significant centripetal acceleration occurs [8]. The combination with the aforementioned NLM and ERM allows to exploit attitude estimation with the EKF maintaining the reduced state vector consisting only of position and velocity estimates or corrections, respectively.

ROS is a cross-platform middle-ware that uses the publish–subscribe messaging system for data exchange and communication over TCP sockets between processes (ROS nodes), regardless the machine they are running on. Therefore, the real-time implementation consisted of a single C++ node subscribing to wheel encoders and IMU readings (published asynchronously over a CAN bus), computing the NLM+CF algorithm, and publishing the navigation data for further utilization. The internal ROS time-stamps were used for proper process synchronization and an OpenCV Matx template class for appropriate matrix management and operations [28]. The average EKF cycle duration of final real-time implementation of the NLM+CF approach onboard the UGV (Intel Core2Quad CPU Q9100 2.26 GHz, 8-GB RAM) was approximately 40 μs.

V. Conclusion

This paper has presented a comparison and experimental evaluation of four different EKF-based estimation architectures for dead reckoning of a mobile robot providing data from IMU and wheel encoders. The dead reckoning was realized using the EKF with the NLM and the ERM, both with and without a CF for attitude estimation. Field testing in indoor and outdoor environment was carried out to select the best candidate in terms of the navigation performance and the computational load. According to the results, the NLM with a CF was selected as the best approach, running two times faster than the ERM-based approaches, and reaching 0.8 m RMSE and average of 4% RPE of the distance driven. It was then implemented in C++ for onboard processing in ROS. With this algorithm implemented, the robot was then deployed in a real USAR mission in Mirandola, Italy (for further details see [2]).

References

[1] G. J. M. Kruijff, M. Janicek, S. Keshavdas, B. Larochelle, H. Zender, N. J. J. M. Smets, T. Mioch, M. A. Neerincx, J. van Diggelen, F. Colas, M. Liu, F. Pomerleau, R. Siegwart, V. Hlavac, T. Svoboda, T. Petricek, M. Reinstein, K. Zimmermann, F. Pirri, M. Gianni, P. Papadakis, A. Sinha, P. Balmer, N. Tomatis, R. Worst, T. Linder, H. Surmann, V. Tretyakov, S. Corrao, S. Pratzler-Wanczura, and M. Sulk, "Experience in system design for human-robot teaming in urban search & rescue," in *Proc. 8th Int. Conf. Field Service Robot.*, 2014, pp. 111–125.

[2] G. J. M. Kruijff, V. Tretyakov, T. Linder, F. Pirri, M. Gianni, P. Papadakis, M. Pizzoli, A. Sinha, E. Pianese, S. Corrao, F. Priori, S. Febrini, and S. Angeletti, "Rescue robots at earthquake-hit Mirandola, Italy: A field report," in *Proc. IEEE 10th Int. Symp. Safety Security Rescue Robot.*, 2012, pp. 1–8.

[3] B. Mobedi and G. Nejat, "3-D active sensing in time-critical urban search and rescue missions," *IEEE/ASME Trans. Mechatronics*, vol. 17, no. 6, pp. 1111–1119, Dec. 2012.

[4] 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 *Proc. IEEE Int. Conf. Robot. Autom.*, 2012, pp. 619–626.

[5] A. Arsenault, S. A. Velinsky, and T. A. Lasky, "A low-cost sensor array and test platform for automated roadside mowing," *IEEE/ASME Trans. Mechatronics*, vol. 16, no. 3, pp. 592–597, Jun. 2011.

[6] A. Nemra and N. Aouf, "Robust INS/GPS sensor fusion for UAV localization using SDRE nonlinear filtering," *IEEE Sens. J.*, vol. 10, no. 4, pp. 789–798, Apr. 2010.

[7] E. H. Shin, "Estimation techniques for low-cost inertial navigation," Ph.D. dissertation, Dept. Geom. Eng., Univ. Calgary, Calgary, CA, 2005.

[8] V. Kubelka and M. Reinstein, "Complementary filtering approach to orientation estimation using inertial sensors only," in *Proc. IEEE Int. Conf. Robot. Autom.*, 2012, pp. 599–605.

[9] J. F. Vasconcelos, B. Cardeira, C. Silvestre, P. Oliveira, and P. Batista, "Discrete-time complementary filters for attitude and position estimation: Design, analysis and experimental validation," *IEEE Trans. Control Syst. Technol.*, vol. 19, no. 1, pp. 181–198, Jan. 2011.

[10] D. H. Titterton and J. L. Weston, *Strapdown Intertial Navigation Technology*. Lavenham, U.K.: Lavenham Press, 1997.

[11] M. Reinstein and M. Hoffmann, "Dead reckoning in a dynamic quadruped robot based on multimodal proprioceptive sensory information," *IEEE Trans. Robot.*, vol. 29, no. 2, pp. 563–571, Apr. 2013.

[12] G. Cook, *Mobile Robots: Navigation, Control and Remote Sensing*. Hoboken, NJ, USA: Wiley-IEEE Press, 2011.

[13] J. Yi, J. Zhang, D. Song, and S. Jayasuriya, "IMU-based localization and slip estimation for skid-steered mobile robots," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2007, pp. 2845–2850.

[14] D. Endo, Y. Okada, K. Nagatani, and K. Yoshida, "Path following control for tracked vehicles based on slip-compensating odometry," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2007, pp. 2871–2876.

[15] G. Dissanayake and S. Sukkarieh, "The aiding of a low-cost strapdown inertial measurement unit using vehicle model constraints for land vehicle applications," *IEEE Trans. Robot. Autom.*, vol. 17, no. 5, pp. 731–747, Oct. 2001.

[16] G. Galben, "New three-dimensional velocity motion model and composite odometry-inertial motion model for local autonomous navigation," *IEEE Trans. Veh. Technol.*, vol. 60, no. 3, pp. 771–781, Mar. 2011.

[17] M. Reinstein, V. Kubelka, and K. Zimmermann, "Terrain adaptive odometry for mobile skid-steer robots," in *Proc. IEEE Int. Conf. Robot. Autom.*, 2013, pp. 4706–4711.

[18] F. Aghili and A. Salerno, "Driftless 3-D attitude determination and positioning of mobile robots by integration of IMU with two RTK GPSs," *IEEE/ASME Trans. Mechatronics*, vol. 18, no. 1, pp. 21–31, Feb. 2013.

[19] X. Song, L. D. Seneviratne, and K. Althoefer, "A Kalman filter-integrated optical flow method for velocity sensing of mobile robots," *IEEE/ASME Trans. Mechatronics*, vol. 16, no. 3, pp. 551–563, Jun. 2011.

[20] J. Shen, D. Tick, and N. Gans, "Localization through fusion of discrete and continuous epipolar geometry with wheel and IMU odometry," in *Proc. Amer. Control Conf.*, 2011, pp. 1292–1298.

[21] Y. Morales, T. Tsubouchi, and S. Yuta, "Vehicle 3D localization in mountainous woodland environments," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2009, pp. 3588–3594.

[22] WillowGarage. (2012). ROS. [Online]. Available: http://ros.org

[23] M. S. Grewal and A. P. Andrews, *Kalman Filtering: Theory and Practice Using MATLAB*, 2nd ed. Hoboken, NJ, USA: Wiley-IEEE Press, 2001.

[24] R. M. Rogers, *Applied Mathematics in Integrated Navigation Systems*, Reston, VA, USA: AIAA, 2003.

[25] J. A. Farrell, *Aided Navigation: GPS with High Rate Sensors*. New York, NY, USA: McGraw-Hill, 2008.

[26] S. Weiss, M. W. Achtelik, M. Chli, and R. Siegwart, "Versatile distributed pose estimation and sensor self-calibration for an autonomous MAV," in *Proc. IEEE Int. Conf. Robot. Autom.*, 2012, pp. 31–38.

[27] Y. Bar Shalom, X. R. Li, and T. Kirubarajan, *Estimation with Applications to Tracking and Navigation*. Hoboken, NJ, USA: Wiley-Interscience, 2001.

[28] WillowGarage. (2012). OpenCV. [Online]. Available: http://opencv.org

4. Robust data fusion of multimodal sensory information for mobile robots

This section presents our second publication Data Fusion [A.2]. It is a 27-page-long paper published in Journal of Field Robotics. A brief overview of the paper follows:

The **introduction** section states our contributions of this publication and describes specific conditions of search and rescue missions we aim for. The experimental evaluation of proposed system is stressed, especially the analysis of possible fail cases.The **related work** is presented in a separate section explaining difference between proprioceptive and exteroceptive sensors and providing references to works from respective fields. Proprioceptive sensory modalities available on the robot are inertial navigation and track odometry, exteroceptive modalities are visual odometry and laser scan matching localization.

The **system description** introduces the hardware of the robot in detail, mainly its on-board computer and sensor suite (inertial measurement unit, track velocity encoders, omni-directional camera and laser range-finder). Sensory modalities are described in this section as well. They constitute are a layer between raw sensor data and our proposed fusion system, which expects processed measurements in form of velocities, angular rates and possibly acceleration.

The next section describes the **multimodal data fusion** system we propose, first explaining the concept of error state followed by kinematic model used to propagate system state in time. The **measurement error model** then links the system state and error state with measurements. This relation is presented for each sensory modality. A special attention is paid to laser scan matching modality, whose low sampling rate requires slightly different approach.

Thorough **experimental evaluation** has been designed considering the search and rescue mission conditions. The experimental dataset, which is publicly available, is intended to test our approach to localization but also offers to test different approaches, since all sensor data have been recorded as well as accurate reference. The section presents metrics used to evaluate the localization system, shows typical experiments and resulting performance. The overall accuracy is provided separately for indoor and outdoor experiments due to different ways of capturing reference position and attitude. The important part of the evaluation is the fail case analysis, which focuses on possible modality failures and examines localization system behavior in these cases. Our finding are then summed up in the **conclusion** section.

Robust Data Fusion of Multimodal Sensory Information for Mobile Robots

Vladimír Kubelka
Center for Machine Perception, Dept. of Cybernetics, Faculty of Electrical Engineering, Czech Technical University in Prague, Technicka 2, 166 27, Prague 6, Czech Republic
e-mail: kubelka.vladimir@fel.cvut.cz
Lorenz Oswald, François Pomerleau, and Francis Colas
ETH Zurich, Tannenstrasse 3, 8092, Zurich, Switzerland
e-mail: loswald@student.ethz.ch, francois.pomerleau@mavt.ethz.ch, francis.colas@mavt.ethz.ch
Tomáš Svoboda and Michal Reinstein
Center for Machine Perception, Dept. of Cybernetics, Faculty of Electrical Engineering, Czech Technical University in Prague, Technicka 2, 166 27, Prague 6, Czech Republic
e-mail: svobodat@fel.cvut.cz, reinstein.michal@fel.cvut.cz

Urban search and rescue (USAR) missions for mobile robots require reliable state estimation systems resilient to conditions given by the dynamically changing environment. We design and evaluate a data fusion system for localization of a mobile skid-steer robot intended for USAR missions. We exploit a rich sensor suite including both proprioceptive (inertial measurement unit and tracks odometry) and exteroceptive sensors (omnidirectional camera and rotating laser rangefinder). To cope with the specificities of each sensing modality (such as significantly differing sampling frequencies), we introduce a novel fusion scheme based on an extended Kalman filter for six degree of freedom orientation and position estimation. We demonstrate the performance on field tests of more than 4.4 km driven under standard USAR conditions. Part of our datasets include ground truth positioning, indoor with a Vicon motion capture system and outdoor with a Leica theodolite tracker. The overall median accuracy of localization—achieved by combining all four modalities—was 1.2% and 1.4% of the total distance traveled for indoor and outdoor environments, respectively. To identify the true limits of the proposed data fusion, we propose and employ a novel experimental evaluation procedure based on failure case scenarios. In this way, we address the common issues such as slippage, reduced camera field of view, and limited laser rangefinder range, together with moving obstacles spoiling the metric map. We believe such a characterization of the failure cases is a first step toward identifying the behavior of state estimation under such conditions. We release all our datasets to the robotics community for possible benchmarking.

1. INTRODUCTION

Mobile robots are sought for many tasks, from tour-guide robots to autonomous cars. With the rapid advance in sensor technology, it has been possible to embed richer sensor suites and extend the perception capabilities. Such sensor suites provide multimodal information that naturally ensures perception robustness, allowing also better means of self-calibration, fault detection, and recovery—given that appropriate data fusion methods are exploited. Independently from the application, a key issue of mobile robotics is state estimation. It is crucial for both perception, such as mapping, and action, such as avoiding obstacles or terrain adaptation.

In this paper, we address the problem of data fusion for localization of an unmanned ground vehicle (UGV) intended for urban search and rescue (USAR) missions. There has been a significant effort presented in the field of USAR for robot localization that mostly aims for a minimal suitable sensing setup, usually exploiting the inertial measurements aided by either vision or laser data. Having sufficient onboard computational power, we therefore aim for a richer sensor suite and hence better robustness and reliability. Therefore, our UGV used in this work (see Figure 1) embeds track encoders, an inertial measurement unit (IMU), an omnidirectional camera, and a rotating laser rangefinder.

Our first contribution lies in the development of a model for such multimodal data fusion using an extended Kalman filter (EKF), especially in the way we incorporate sensors with slow and fast measurement update rates. To cope with such a significant difference in the update rates of various sensor modalities, we concentrated the model design on integrating the slow laser and visual odometry with the faster IMU and track odometry measurements. For

Journal of Field Robotics 32(4), 447–473 (2015)
View this article online at wileyonlinelibrary.com • DOI: 10.1002/rob.21535

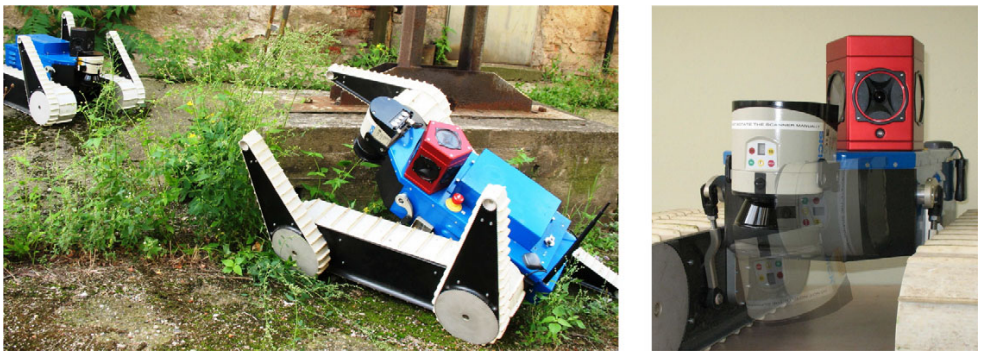


Figure 1. Picture of two USAR UGVs used for experimental evaluation (FP7-ICT-247870 NIFTi project) and a detail of the sensor setup (a PointGrey Ladybug 3 omnicamera and a rotating SICK LMS-151 laser rangefinder). See Section 3.1 for more details.

this purpose, we propose and investigate three different possible methods—one of them, the *trajectory approach* (see Section 4.3.3 for further details), is our contribution that we compare to the *velocity approach*, which is a common state-of-the-art practice. We show that a standard EKF designed with the *velocity approach* does not cope well with such significant differences in the frequency, whether or not our proposed *trajectory approach* does.

The context of USAR missions implicitly defines the challenges and limitations of our application. The environment is often unstructured (collapsed buildings) and unstable (moving objects or other ongoing changes, deformable terrain causing high slippage). Robots need to cope with indoor-outdoor transitions (change from confined to open spaces), as well as bad lighting conditions with rapid changes and sometimes decreased visibility (smoke and fire). These are essentially the main challenges that come with the sensor data we process. Therefore, our main contribution lies in the actual experimental evaluation and analysis of the limits of the proposed filter. We review the different sensing modalities and their expected failure cases to assess the impact of possible data degradation (or outage) on the overall precision of localization. We believe that the field deployment of state estimation for multimodal data fusion needs to be characterized both under standard expected conditions and for partial or full failures of sensing modalities. Indeed, robustness to sensor data outage or degradation is a key element to the scaling up of a field robotics system. Therefore, we evaluate our filter using several hours and kilometers of experimental data validated by indoor or outdoor ground truth measurements. To share this contribution with the robotics community, we release all the captured datasets (including the ground truth measurements) to be used as benchmarks.[1]

[1]The datasets are available as *bagfiles* for ROS at https://sites.google.com/site/kubelvla/public-datasets

The state of the art of sensor fusion for state estimation is elaborated in Section 2. In Section 3, we present the hardware and software used in this work before describing in detail the design of our data fusion algorithm (Section 4). In Section 5, we explain our experimental evaluation including our fail-case methodology before a discussion and conclusion (Section 6).

2. RELATED WORK

In general, the information obtained from various sensors can be classified as either proprioceptive (inertial measurements, joint sensors, motor or wheel encoders, etc.) or exteroceptive [global positioning system (GPS), cameras, laser rangefinder, ultrasonic sensors, magnetic compass, etc.]. Exteroceptive sensors that acquire information from the environment can also be used to perceive external landmarks that are necessary for long-term precision in navigation tasks. In modern mobile robots, a popular solution lies usually in the combination of a proprioceptive component in the form of an inertial navigation system (INS) (Titterton and Weston, 1997) that captures the body dynamics at high frequency, and an external source of aiding, using vision (Chowdhary, Johnson, Magree, Wu, & Shein, 2013) or range measurements (Bachrach, Prentice, He, & Roy, 2011). The key issue lies in the appropriate integration of the different characteristics of the different sensor modalities.

As was repeatedly shown, the combination of an IMU with wheel odometry is a popular technique to localize a mobile robot in a dead-reckoning manner. It generally allows for a very high sampling frequency as well as processing rate, usually without excessive computational load. Dead reckoning can be used for short-term navigation without any necessity of perceiving the surrounding environment via exteroceptive sensors. In real outdoor conditions, the dynamically changing environment often causes signal degradation or even outage of exteroceptive sensors. However, proprioceptive sensing, in principle, is too prone

to accumulating errors to be used as a stand-alone solution. Computational and environmental errors as well as errors caused by misalignment and instrumentation cause the dead-reckoning system to drift quickly with time. Moreover, motor encoders do not reflect the true path, especially the heading of the vehicle, in the case of frequent wheel slip. In Yi, Zhang, Song, and Jayasuriya (2007) and Anousaki and Kyriakopoulos (2004), an improvement through the skid-steer model of a four-wheel robot is presented, based on a Kalman filter estimating trajectory using velocity constraints and slip estimate. An alternative method appears in Endo, Okada, Nagatani, and Yoshida (2007), where the IMU and odometry are used to improve tracked vehicle navigation via slippage estimates. We addressed this problem in Reinstein, Kubelka, and Zimmermann (2013). Substantial effort has also been made to investigate the odometry derived constraints (Dissanayake, Sukkarieh, Nebot, & Durrant-Whyte, 2001) or innovation of the motion models (Galben, 2011). Concerning all the references so far, localization of the navigated object via dead reckoning was performed only in two dimensions. There exist solutions providing real three-dimensional (3D) odometry derived from the rover-type multiwheel vehicle design (Lamon & Siegwart, 2004). Nevertheless, the error is still about one order of magnitude higher than what we aim to achieve (below 2% of the total distance traveled).

However, if long-term precision and reliability are to be guaranteed, dead-reckoning solutions require other exteroceptive aiding sensor systems. In the work of Shen, Tick, and Gans (2011), it is shown that a very low-cost IMU and odometry dead-reckoning system can be realized and successfully combined with visual odometry (VO) (Sakai, Tamura, & Kuroda, 2009; Scaramuzza & Fraundorfer, 2011) to produce a reliable navigation system. With the increasing onboard computational power, visual odometry is becoming very popular even for large-scale outdoor environments. Most solutions are based on the EKF (Chowdhary, Johnson, Magree, Wu, & Shein, 2013; Civera, Grasa, Davison, & Montiel, 2010; Konolige, Agrawal, & Sola, 2011; Oskiper, Chiu, Zhu, Samarasekera, & Kumar, 2010) or a dimensional-bounded EKF with a landmark classifier introduced in Jesus and Ventura (2012). However, in Rodriguez F, Fremont, and Bonnifait (2009) it is pointed out that a tradeoff between precision and execution time has to be examined. Moreover, VO degrades due to high rotational speed movements and it is susceptible to illumination changes and lack of sufficient scene texture (Scaramuzza & Fraundorfer, 2011).

Another typically used six degree of freedom (6 DOF) aiding source is a laser rangefinder, which is used for estimating vehicle motion by matching consecutive laser scans and creating a 3D metric map of the environment (Suzuki, Kitamura, Amano, & Hashizume, 2010; Yoshida, Irie, Koyanagi, & Tomono, 2010). Examples of successful application can be found for both indoor use—without IMU but combined with vision (Ellekilde, Huang, Miro, & Dissanayake, 2007)—as well as outdoor use—relying on the IMU (Bachrach et al., 2011). As in case of the visual odometry, solutions using EKF are often proposed (Bachrach, Prentice, He, & Roy, 2011; Morales, Carballo, Takeuchi, Aburadani, & Tsubouchi, 2009). The most popular approach of scan matching is based on the iterative closest point (ICP) algorithm first proposed by Besl and McKay (1992) and in parallel by Chen and Medioni (1991). More recently, Nuchter, Lingemann, Hertzberg, and Surmann (2007) proposed a 6D simultaneous localization and mapping (SLAM) system relying mainly on ICP. Closer to USAR applications, Nagatani et al. (2011) demonstrated the use of ICP in exploration missions and used a pose graph minimization scheme to handle multirobot mapping. Kohlbrecher, Stryk, Meyer, and Klingauf (2011) proposed a localization system combining a 2D laser SLAM with a 3D IMU/odometry-based navigation subsystem. A combination of 3D-landmark-based SLAM and multiple proprioceptive sensors is also presented in Chiu, Williams, Dellaert, Samarasekera, and Kumar (2013), whose work focuses mainly on a low latency solution while estimating the navigation state by means of a sliding-window factor graph. The problem of utilizing several sensors for localization that may provide contradictory measurements is discussed in Sukumar, Bozdogan, Page, Koschan, & Abidi (2007). The authors use Bayes filters to estimate sensor measurement uncertainty and sensor validity to intelligently choose a subset of sensors that contribute to localization accuracy. As opposed to the later publications realized in the context of SLAM, we only consider the results of the ICP algorithm as a local pose measurement, similarly to Almeida and Santos (2013), who use the ICP algorithm to extract the steering angle and linear velocity of a carlike vehicle to update its nonholonomic model of motion. In our approach, the 3D reconstruction of the environment is considered locally coherent, and neither loop detection nor error propagation is used.

As stated in Kelly, Sibley, Barfoot, & Newman (2012), it is the right time to address issues concerning the state of the art in long-term navigation and autonomy. In this respect, the benefits and challenges of repeatable long-range driving were addressed in Barfoot, Stenning, Furgale, and McManus (2012). In this context, we believe that bringing more insight into multimodality state estimation algorithms is an important step for the long-term stability of a USAR system evolving in a complex range of environments.

Regarding multimodal data fusion, we built on our previous work concerning complementary filtering (Kubelka & Reinstein, 2012), odometry modeling (Reinstein et al., 2013), and design of EKF error models (Reinstein & Hoffmann, 2013), even though the latter work applied to a legged robot.

3. SYSTEM DESCRIPTION

Our system is aimed at high state estimation accuracy while ensuring robust performance against rough terrain

navigation and obstacle traversals. We selected four modalities to achieve this goal: the inertial measurements (IMU), odometry data (OD), visual odometry (VO), and laser rangefinder data (ICP) processed by the ICP algorithm. This section explains the motion capabilities of the Search & Rescue platform and the preprocessing computation applied to its sensors in order to extract meaningful inputs for the state estimation. These explanations provide a motivation for a list of states to be estimated by the EKF described in Section 4.

3.1. Mobile Robotic Platform

Figure 1 presents the UGV designed for the USAR mission that we use in this paper. As described in Kruijff et al. (2012), this platform was deployed multiple times in collaboration with various rescue services (Fire Department of Dortmund/Germany, Vigili del Fuoco/Italy). It has two bogies linked by a differential that allows a passive adaptation to the terrain. On each of the tracks, there are two independent flippers that can be position-controlled in order to increase the mobility in difficult terrain. For example, they can be unfolded to increase the support polygon, which helps to overcome gaps and increase stability on slopes. They can also be raised to help with climbing over higher obstacles. Given that the robot was designed to operate in 3D unstructured environments, the state estimation system needs to provide a 6 DOF localization.

Encoders are placed on the differential, giving the angle between the two bogies and the body, on the tracks to give their current velocity, and on each flipper to give its position with respect to its bogies. Inside the body, vertical to the center of the robot, lies the Xsens MTi-G IMU providing angular velocities and linear acceleration along each of the three axes. The IMU data capture the body dynamics at the high rate of 90 Hz. GPS is not taken into account due to the low availability of the signal indoors or in close proximity with buildings. The magnetic compass is also easily disturbed by metallic masses, pipes, and wires, which make it highly unreliable, and hence we do not use it.

The exteroceptive sensors of the robot consist of an omnidirectional camera and a laser rangefinder. The omnidirectional camera is the PointGrey Ladybug 3 and produces a 12 megapixels stitched omnidirectional images at 5–6 Hz. The omnidirectionality of the sensor provides a stronger stability of rotation estimation at the expense of scale estimation, which would be better handled by a stereocamera. The laser rangefinder used is the Sick LMS-151 mounted on a rolling axis in front of the robot. The laser spins left and right alternately, taking a full 360° scan at approximately 0.3 Hz to create a point cloud of around 55,000 points.

3.2. Inertial Data Processing

Although the precision and reliability of the IMU measurements is sufficient in the short term, in the long term the information provided suffers from random drift that, together with integrated noise, causes unbounded error growth. To cope with these errors, all the six sensor biases have to be estimated (see Section 4.1 for more details). Therefore, we have included sensor biases in the state space of the proposed EKF estimator. Furthermore, correct calibration of the IMU output and its alignment with respect to the robot's body frame has to be assured.

3.3. Odometry for Skid-steer Robots

Our platform is equipped with caterpillar tracks, and therefore steering is realized by setting different velocities for each of the tracks (*skid-steering*). The encoders embedded in the tracks of the platform measure the left and right track velocities at approximatively 15 Hz. However, in contradistinction to differential robots, the odometry for skid-steering vehicles has significant uncertainties. Indeed, as soon as there is a rotation, the tracks must either deform or slip significantly. The slippage is affected by many parameters including the type and local properties of the terrain. To keep the computation complexity low, we assume only a simple odometry model and we do not model the slippage. Instead, we take advantage of the exteroceptive modalities in our data fusion to observe the true motion dynamics using different sources of information. Hence, the fusion compensates for cases in which the tracks are slipping because the surface is slippery or because of an obstacle blocking the robot. Another advantage of using caterpillar tracks odometry lies in the opportunity to exploit nonholonomic constraints. Further explanations on those constraints are given in Section 4.3.

3.4. ICP-based Localization

Using as *Input* the current 3D point cloud, a registration process is used to estimate the pose of the robot with respect to a global representation called *Map*. We used a derivation of the point-to-point ICP algorithm introduced by Chen and Medioni (1991) combined with the trimmed outlier rejection presented by Chetverikov, Svirko, Stepanov, and Krsek (2002).

The implementation uses `libpointmatcher`,[2] an open-source library fast enough to handle real-time processing while offering modularity to cover multiple scenarios as demonstrated in Pomerleau, Colas, Siegwart, & Magnenat (2013). The complete list of modules used with their main parameters can be found in Table I. More specifically, the configuration of the rotating laser produced a high density of points in front of the robot, which was desirable to predict collision but not beneficial to the registration minimization. Thus, we forced the maximal density to 100 points per m^3 after having randomly subsampled the point cloud

[2]https://github.com/ethz-asl/libpointmatcher

Table I. Configurations of ICP chains for the NIFTi mapping applications.

	Step	*Module*	*Description*
Input	Read. filtering	`SimpleSensorNoise`	SickLMS
		`SamplingSurfaceNormal`	keep 80%, surface normals based on 20 NN
		`ObservationDirection`	add vector pointing toward the laser
		`OrientNormals`	orient surface normals toward the obs. direction
		`MaxDensity`	subsample to keep point with density of 100 pts/m^3
Registration	Ref. filtering	-	processing from the rows Map
	Read. filtering	-	processing from the rows Input
	Data association	`KDTree`	kd-tree matching with 0.5 m max. distance, $\epsilon = 3.16$
	Outlier filtering	`TrimmedDist`	keep 80% closest points
		`SurfaceNormal`	remove paired normals angle $> 50^\circ$
	Error min.	`PointToPlane`	point-to-plane
	Trans. checking	`Differential`	min. error below 0.01 m and 0.001 rad
		`Counter`	iteration count reached 40
		`Bound`	transformation fails beyond 5.0 m and 0.8 rad
Map	Ref. filtering	`SurfaceNormal`	Update normal and density, 20 NN, $\epsilon = 3.16$
		`MaxDensity`	subsample to keep point with density of 100 pts/m^3
		`MaxPointCount`	subsample 70% if more than 600,000 points

in order to finish the registration and the map maintenance within 2 s. We expected the error on prealignment of the 3D scans to be less than 0.5 m based on the velocity of the platform and the number of ICPs per second that were to be executed. So we used this value to limit the matching distance. We also removed paired points with an angle difference larger than 50° to avoid the reconstruction of both sides of walls from collapsing when the robot was exploring different rooms. The surface normal vector used for the *outlier filtering* and for the *error minimization* are computed using 20 nearest neighbors (NNs) of every point within a single point cloud. As for the global map, we maintained a density of 100 points per m^3 every time a new input scan was merged in it. A maximum of 600,000 points were kept in memory to avoid degradation of the computation time when exploring a larger environment than expected. However, the only output of the ICP algorithm we consider is the robot's localization, i.e., position and orientation relative to its inner 3D point-cloud map. We do not attempt to create a globally consistent map and we do not exploit the map in any other way than for analysis of the ICP performance (no map corrections or loop closures are performed).

There is one ICP-related issue observed with our platform. Although the ICP creates a locally precise metric map, the map as a whole tends to slightly twist or bend (we do not perform any loop-closure). This is why the position and the attitude estimated by the ICP odometry collide with other position information sources. Another limitation is the refresh rate of the pose measurements limited to 0.3 Hz. This rate is far from our fastest measurement (i.e., the IMU at 90 Hz), which poses a linearization problem. For these reasons, we investigated three different types of measurement models; see Section 4.3.3 for details.

Furthermore, the true bottleneck of the ICP-based localization lies in the way it is realized on our platform and hence is prone to mechanical issues. As the laser rangefinder has to be turning to provide a full 3D point cloud, in an environment with high vegetation such a mechanism is easily struck, causing this modality to fail. Large open spaces, indoor/outdoor transitions, or significantly large moving obstacles can also cause the ICP to fail updating the metric map. Since this modality is very important, we analyzed these failure cases in Section 5.4.

3.5. Visual Odometry

Our implementation of visual odometry generally follows the usual scheme (Scaramuzza & Fraundorfer, 2011; Tardif, Pavlidis, & Daniilidis, 2008). The VO computation runs solely on the robot onboard computer and estimates the pose at the frame rate 2–3 Hz, which, compared to the robot speed, is sufficient. It does search for correspondences (i.e., image matching) (Rublee, Rabaud, Konolige, & Bradski, 2011), landmark reconstruction, and sliding bundle adjustment (Fraundorfer & Scaramuzza, 2012; Kummerle, Grisetti, Strasdat, Konolige, & Burgard, 2011), which refines the landmark 3D positions and the robot poses. The performance essentially depends on the visibility and variety of landmarks. The more variant landmarks are visible at more positions, the more stable and precise is the pose estimation. The process uses panoramic images constructed from spherical approximation of the Ladybug camera model. The Ladybug camera is approximated as one central camera. The error of the approximation is acceptable for landmarks that are a few meters from the robot.

The visual odometry starts with detecting and matching features in two consecutive images. We use OpenCV implementation of the Orb keypoint detector and descriptor (Rublee et al., 2011). Only the matches that are distinctive above a certain threshold survive. The initial matching is supported by a guided matching that uses an initial estimate of the robot movement. The robot movement is estimated by the five-point solver (Li & Hartley, 2006) encapsulated in RANSAC iterations. As the error measure, we use the angular deviation of points from epipolar planes. This is less precise than the usual distance from epipolar lines. However, as we work with spherical projection, we have epipolar curves. Computing angular deviations is faster than computing the distance to the epipolar curve. The movement estimate projects already known landmarks, and we can actively search around the projection. The feature tracks are updated and associated with landmarks if they pass an observation consistency test. The landmark 3D position is triangulated from all possible observations, and the complete estimate of landmark and robot positions is refined by a bundle adjustment (Kummerle et al., 2011).

Using an almost omnidirectional camera for the robot motion estimation is geometrically advantageous (Brodsky, Fermueller, & Aloimonos, 1998; Svoboda, Pajdla, & Hlaváč, 1998). The scale estimation however, depends on the precision of 3D reconstruction where the omnidirectionality does not really help. It is also important to note that the omnidirectional camera we use sits very low above the terrain (below 0.5 m) and directly on the robot body. This makes a huge difference compared to, e.g., Tardif et al. (2008), where the camera is more than 2 m above the terrain and sees the ground plane much better than our camera. Estimation of the yaw angle is still well conditioned since it relies mostly on the side correspondences. The pitch estimation, however, would sometimes need more landmarks on the ground plane. The pitch part of the motion induces the largest disparity of the correspondences in the front and back cameras. Unfortunately, the back view is significantly occluded by the battery cover. This is especially problematic in the street scenes where the robot moves along the street; see, e.g., Figure 11. The front cameras see the street level better; however, the uniform texture of the tar surface often generates only a few reliable correspondences. The search for correspondences is further complicated by the tilting flippers, which occlude the field of view and induce outliers. The second problem is the agility of the robot combined with the relatively low frequency of the visual odometry. The robot can turn on a spot very quickly, much quicker than an ordinary wheeled car. Even worse, the quick turn is the usual way in which the movement direction is changed. This makes correspondence search difficult. In the future versions of visual odometry, we want to improve the landmark management in order to resolve the problem of too few landmarks surviving the sudden turn. We also think about replacing the approximate spherical model by reformulating it in a multiview model.

4. MULTIMODAL DATA FUSION

The core of the data fusion system is realized by an error-state EKF inspired by the work of Weiss (2012). The description of the multimodal data fusion solution we propose can be divided into two parts. First is the process error model for the EKF, which shows how we model the errors, which we aim to estimate and use for corrections. The second part is the measurement model, which couples the sensory data coming at different rates.

The overall scheme of our proposed approach is shown in Figure 2. Raw sensor data are preprocessed and used as measurements in the error state EKF (the *FUSION* block). There is no measurement rejection implemented; based on the assumption that fusion of several sensor modalities should deal with anomalous data inherently—for details see Sections 5 and 6—this, however, will be subjected to a future work. As is apparent from Figure 2, measurement rates significantly differ among the sensor modalities—the main difference is especially between the IMU at 90 Hz and the ICP output at 0.3 Hz. Having the update rate of the EKF at 90 Hz, the experiments have proven that this issue is crucial and has to be resolved as part of the filter design to ensure reliable output from the fusion process (see Section 5.3.3). In our case, this problem concerns mainly the ICP-based localization that provides measurements at a very low rate of 0.3 Hz—too low to capture the motion dynamics as the IMU does (i.e., the motion dynamics spectrum gets subsampled). During these 3 s, real-world disturbances (which are often non-Gaussian and difficult to model and predict, e.g., tracks slippage) accumulate. This was the motivation to investigate various ways of fusing measurements at significantly different rates. Three proposed approaches that incorporate the ICP measurements are described in Section 4.3.3.

4.1. Process Error Model

For the purpose of localization, we model our robot as a rigid body with constant angular rate and constant rate of change of velocity ($\dot{\omega} = 0, \dot{v} = \text{const}$). The presence of constant gravitational acceleration is expected and incorporated into the system model; no dissipative forces are considered.

We define four coordinate frames: the *R(obot)* frame coincides with the center of the robot, the *I(MU)* frame represents the inertial measurement unit coordinate frame as defined by the manufacturer, the *O(dometry)* frame represents the tracked gear-frame, and the *N(avigation)* frame represents the world frame. In all these frames, the North-West-Up axes convention is followed, with the x-axis pointing forward (or to the North in the N-frame), the y-axis pointing to the left (or to the West), and the z-axis

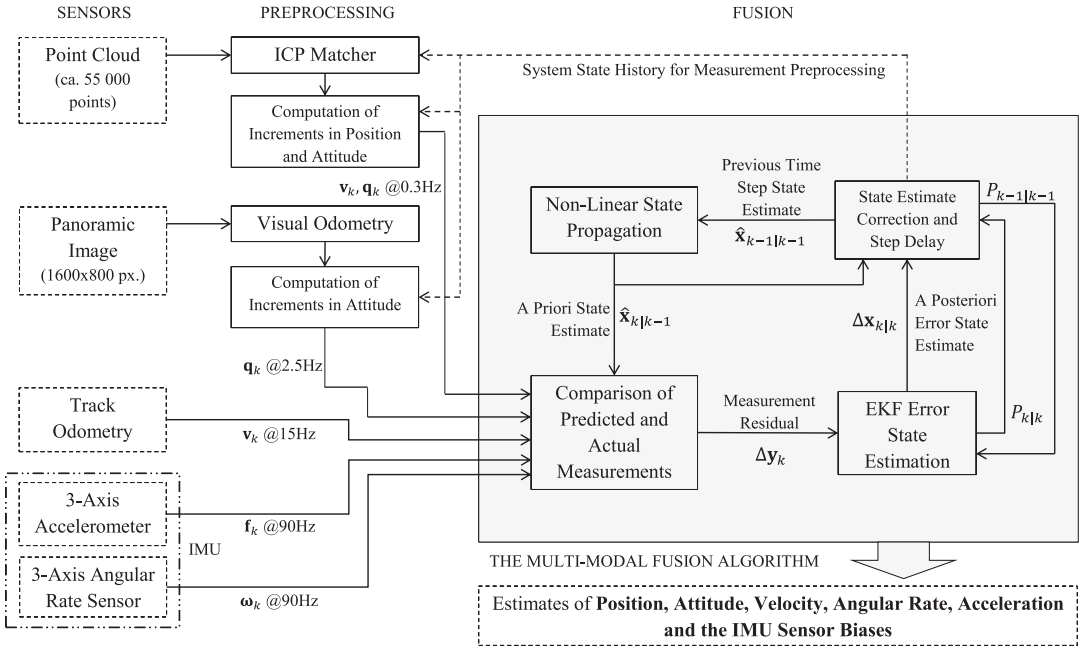


Figure 2. The scheme of the proposed multimodal data fusion system [$\boldsymbol{\omega}$ is angular velocity, $\mathbf{f}$ is specific force (Savage, 1998), $\mathbf{v}$ is velocity, and $\mathbf{q}$ is quaternion representing attitude].

pointing upward. Rotations about each axis follow the *right-hand rule*. The fundamental part of the system design is the differential equations describing the development of the states in time. The state space with the corresponding errors is defined as

$$\mathbf{x} = \begin{bmatrix} \mathbf{p}_N \\ \mathbf{q}_N^R \\ \mathbf{v}_R \\ \boldsymbol{\omega}_R \\ \mathbf{f}_R \\ \mathbf{b}_{\omega,I} \\ \mathbf{b}_{f,I} \end{bmatrix}, \qquad \Delta\mathbf{x} = \begin{bmatrix} \Delta\mathbf{p}_N \\ \delta\boldsymbol{\theta} \\ \Delta\mathbf{v}_R \\ \Delta\boldsymbol{\omega}_R \\ \Delta\mathbf{f}_R \\ \Delta\mathbf{b}_{\omega,I} \\ \Delta\mathbf{b}_{f,I} \end{bmatrix}, \tag{1}$$

where $\mathbf{p}_N$ is position of the robot in the N-frame, $\mathbf{q}_N^R$ is a unit quaternion representing its attitude, $\mathbf{v}_R$ is the velocity expressed in the R-frame, $\boldsymbol{\omega}_R$ is the angular rate, $\mathbf{f}_R$ is the specific force (Savage, 1998), and $\mathbf{b}_{\omega,I}$ and $\mathbf{b}_{f,I}$ are accelerometer and angular rate sensor IMU-specific biases expressed in the I-frame.

The error state $\Delta\mathbf{x}$ is defined—following the idea of Weiss (2012) (Eq. 3.25)—as the difference between the system state and its estimate $\Delta\mathbf{x} = \mathbf{x} - \hat{\mathbf{x}}$ except for attitude, where the rotation error vector $\delta\boldsymbol{\theta}$ is the vector part of the error quaternion $\delta\mathbf{q} = \mathbf{q} \otimes \hat{\mathbf{q}}^{-1}$ multiplied by 2; $\otimes$ represents quaternion multiplication as defined in Breckenridge (1999).

The states and the error states of the robot, modeled as a rigid body movement, propagate in time according to the following equations:

$$\dot{\mathbf{p}}_N = C^T_{(\mathbf{q}_N^R)}\mathbf{v}_R, \qquad \Delta\dot{\mathbf{p}}_N \approx C^T_{(\hat{\mathbf{q}}_N^R)}\Delta\mathbf{v}_R - C^T_{(\hat{\mathbf{q}}_N^R)}\delta\boldsymbol{\theta}, \tag{2}$$

$$\dot{\mathbf{q}}_N^R = \frac{1}{2}\Omega(\boldsymbol{\omega}_R)\mathbf{q}_N^R, \qquad \delta\dot{\boldsymbol{\theta}} \approx -\lfloor\hat{\boldsymbol{\omega}}_R\rfloor\delta\boldsymbol{\theta} + \Delta\boldsymbol{\omega}_R + \mathbf{n}_\theta, \tag{3}$$

$$\begin{aligned} \dot{\mathbf{v}}_R &= \mathbf{f}_R - C_{(\mathbf{q}_N^R)}\mathbf{g}_N + \lfloor\mathbf{v}_R\rfloor\boldsymbol{\omega}_R, \\ \Delta\dot{\mathbf{v}}_R &\approx \Delta\mathbf{f}_R - \lfloor C_{(\hat{\mathbf{q}}_N^R)}\mathbf{g}_N\rfloor\delta\boldsymbol{\theta} + \lfloor\hat{\mathbf{v}}_R\rfloor\Delta\boldsymbol{\omega}_R - \lfloor\hat{\boldsymbol{\omega}}_R\rfloor\Delta\mathbf{v}_R + \mathbf{n}_v, \end{aligned} \tag{4}$$

$$\begin{aligned} \dot{\boldsymbol{\omega}}_R &= 0, \qquad \dot{\mathbf{f}}_R = 0, \qquad \dot{\mathbf{b}}_{\omega,I} = 0, \qquad \dot{\mathbf{b}}_{f,I} = 0, \\ \Delta\dot{\boldsymbol{\omega}}_R &= \mathbf{n}_\omega, \qquad \Delta\dot{\mathbf{f}}_R = \mathbf{n}_f, \\ \Delta\dot{\mathbf{b}}_{\omega,I} &= \mathbf{n}_{b,\omega}, \qquad \Delta\dot{\mathbf{b}}_{f,I} = \mathbf{n}_{b,f}, \end{aligned} \tag{5}$$

where the derivation of the left part of Eq. (3) can be found in Trawny and Roumeliotis (2005) (Eq. 110) and the left part of Eq. (4) is based on Nemra and Aouf (2010) (Eq. 5); the difference from the original is caused by different ways of expressing attitude. The right parts of Eqs. (2)–(4) can be derived by neglecting higher-order error terms and by

an approximation of the error in attitude by the rotation error vector $\delta\boldsymbol{\theta}$ following Weiss (2012) (Eq. 3.44). We define $\mathbf{g}_N = [0, 0, g]^T$, $\mathbf{n}_{(.)}$ are the system noise terms, and $\Omega(\boldsymbol{\omega}_R)$ in Eq. (3) is a matrix representing quaternion and vector product operation (Trawny & Roumeliotis, 2005, Eq. 108). It is constructed as

$$\Omega(\boldsymbol{\omega}) = \begin{bmatrix} 0 & \omega_3 & -\omega_2 & \omega_1 \\ -\omega_3 & 0 & \omega_1 & \omega_2 \\ \omega_2 & -\omega_1 & 0 & \omega_3 \\ -\omega_1 & -\omega_2 & -\omega_3 & 0 \end{bmatrix}. \tag{6}$$

In Eq. (5), time derivations of angular rates and specific forces are equal to zero—usually, they are considered rather as input than state. However, we included them into the state vector to be updated by the EKF. The error model equations can be expressed in compact matrix form:

$$\Delta\dot{x} = F_c \Delta\mathbf{x} + G_c \mathbf{n}, \tag{7}$$

where F_c is a continuous-time state transition matrix, G_c is a noise-coupling matrix, and $\mathbf{n}$ is a noise vector composed of all the $\mathbf{n}_{(.)}$ terms; the F_c matrix is

$$F_c = \left[\begin{array}{c|c|c|c|c|c|c} \emptyset_3 & -C^T_{(\hat{q}^R_N)} & C^T_{(\hat{q}^R_N)} & \emptyset_3 & \emptyset_3 & \emptyset_3 & \emptyset_3 \\ \hline \emptyset_3 & -\lfloor \hat{\omega}_R \rfloor & \emptyset_3 & I_3 & \emptyset_3 & \emptyset_3 & \emptyset_3 \\ \hline \emptyset_3 & -\lfloor C_{(\hat{q}^R_N)} g_N \rfloor & -\lfloor \hat{\omega}_R \rfloor & \lfloor \hat{v}_R \rfloor & I_3 & \emptyset_3 & \emptyset_3 \\ \hline \emptyset_3 & \emptyset_3 & \emptyset_3 & \emptyset_3 & \emptyset_3 & \emptyset_3 & \emptyset_3 \\ \hline \emptyset_3 & \emptyset_3 & \emptyset_3 & \emptyset_3 & \emptyset_3 & \emptyset_3 & \emptyset_3 \\ \hline \emptyset_3 & \emptyset_3 & \emptyset_3 & \emptyset_3 & \emptyset_3 & \emptyset_3 & \emptyset_3 \\ \hline \emptyset_3 & \emptyset_3 & \emptyset_3 & \emptyset_3 & \emptyset_3 & \emptyset_3 & \emptyset_3 \end{array}\right] \tag{8}$$

and the $G_c\mathbf{n}$ term is

$$G_c\mathbf{n} = \left[\begin{array}{c|c|c|c|c|c|c} \emptyset_3 & \emptyset_3 & \emptyset_3 & \emptyset_3 & \emptyset_3 & \emptyset_3 & \emptyset_3 \\ \hline I_3 & \emptyset_3 & \emptyset_3 & \emptyset_3 & \emptyset_3 & \emptyset_3 & \emptyset_3 \\ \hline \emptyset_3 & I_3 & \emptyset_3 & \emptyset_3 & \emptyset_3 & \emptyset_3 & \emptyset_3 \\ \hline \emptyset_3 & \emptyset_3 & I_3 & \emptyset_3 & \emptyset_3 & \emptyset_3 & \emptyset_3 \\ \hline \emptyset_3 & \emptyset_3 & \emptyset_3 & I_3 & \emptyset_3 & \emptyset_3 & \emptyset_3 \\ \hline \emptyset_3 & \emptyset_3 & \emptyset_3 & \emptyset_3 & I_3 & \emptyset_3 & \emptyset_3 \\ \hline \emptyset_3 & \emptyset_3 & \emptyset_3 & \emptyset_3 & \emptyset_3 & I_3 & \emptyset_3 \end{array}\right] \begin{bmatrix} \mathbf{n}_\theta \\ \mathbf{n}_v \\ \mathbf{n}_\omega \\ \mathbf{n}_f \\ \mathbf{n}_{b,\omega} \\ \mathbf{n}_{b,f} \end{bmatrix}. \tag{9}$$

The noise-coupling matrix describes how particular noise terms affect the system state. Each $\mathbf{n}_{(.)}$ term is a random variable with normal probability distribution. The properties of these random variables are described by their covariances in the system noise matrix Q_c. Since they are assumed independent, the matrix Q_c is diagonal, $Q_c = \text{diag}(\sigma^2_{\theta_x}, \sigma^2_{\theta_y}, \sigma^2_{\theta_z}, \sigma^2_{v_x}, \sigma^2_{v_y}, \ldots)$, where σ is the standard deviation.

To implement the proposed model, we have to transform the continuous-time equations to the discrete time domain. We use the Van Loan discretization method (Van Loan, 1978) instead of explicitly expressing the values of the discretized matrices. We substitute into the matrix M defined by Van Loan,

$$M = \begin{bmatrix} -F_c & G Q_c G^T \\ \emptyset & F_c^T \end{bmatrix} \Delta t, \tag{10}$$

and we evaluate the matrix exponential,

$$e^M = \begin{bmatrix} \ddots & F_d^{-1} Q_d \\ \emptyset & F_d^T \end{bmatrix}. \tag{11}$$

The result of the matrix exponential contains the discretized system matrix F_d in the bottom-right part and the discretized system noise matrix Q_d left multiplied by the inversion of F_d in the top-right part. The discretized system matrix F_d can be easily extracted; Q_d can be obtained by left multiplying the upper right part of e^M by F_d.

4.2. State Prediction and Update Using the EKF

The extended Kalman filter (McElhoe, 1966; Smith, Schmidt, & McGee, 1962) is a modification of the Kalman filter (Kalman, 1960), i.e., an optimal observer minimizing the variances of the observed states. Since the error-state EKF is used in our approach, the state of the system is expressed as a sum of the current best estimate ($\hat{\mathbf{x}}$) and some small error ($\Delta\mathbf{x}$). The only difference compared to a standard EKF is that the linearized system matrices F and Q describe only the error state and the error-state covariance propagation in time, rather than the whole state and state covariance propagation in time. This is mainly beneficial from the computational point of view since it simplifies linearization of the system equations. A flow chart describing the error-state EKF computation is shown in Figure 3 and can be decomposed into a series of steps that describe the actual implementation. As new measurements arrive, state estimate ($\hat{\mathbf{x}}$) and its error covariance matrix (P) are available from the previous time-step (or as initialized during first iteration). This state estimate $\hat{\mathbf{x}}$ is propagated in time using the nonlinear system equations. The continuous-time F_c and G_c matrices are evaluated based on the current value of $\hat{\mathbf{x}}$. The Van Loan discretization method is used to obtain discrete forms of F_d and Q_d. Then the error-state covariance matrix P is propagated in time. Expected measurements are compared to the incoming ones, and their difference is expressed in the form of measurement residual $\Delta\mathbf{y}$. Innovation matrix H, expressing the measurement residual as a linear combination of the error-state components, is evaluated. Using the *a priori* estimate of P, H and the variance of the sensor signals expressed as R, the Kalman gain matrix K is computed. The error state $\Delta\mathbf{x}$ is updated using the Kalman gain and the measurement residual; the *a posteriori* estimate of the error-state covariance matrix P is evaluated as well. Finally, the *a priori* state estimate $\hat{\mathbf{x}}$ is corrected using the estimated error $\Delta\mathbf{x}$.

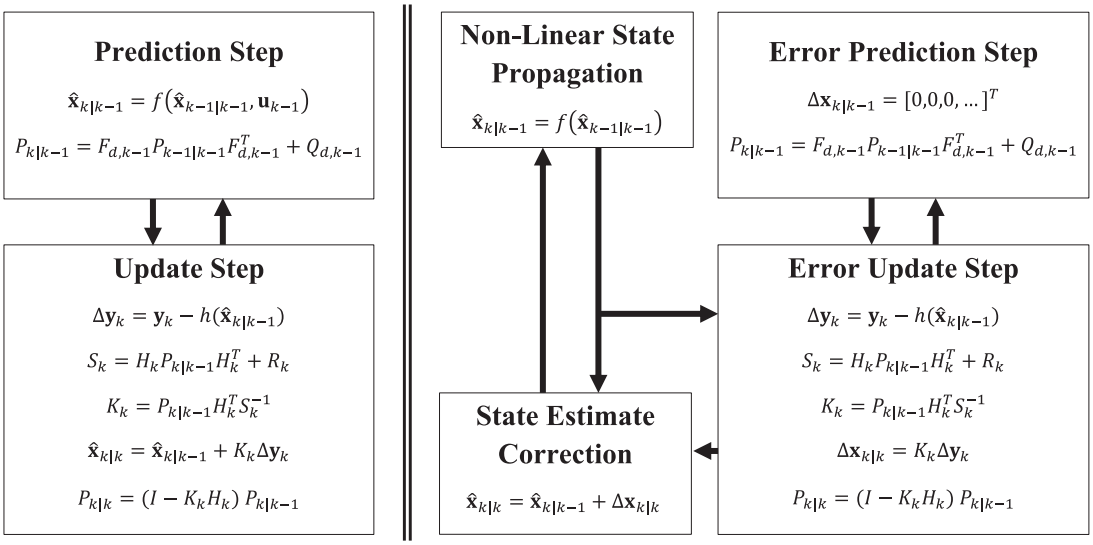


Figure 3. Standard EKF (left) computation flowchart compared to the error state EKF computation flowchart (right): in the error state EKF prediction step, the *a priori* state is estimated using the nonlinear system equation $f()$, and the covariances are estimated using F_d (linearized matrix form of the error state propagation equations). In the update step, the measurement residual $\Delta\mathbf{y}$ is obtained by comparing the incoming measurement $\mathbf{y}$ with its predicted counterpart. The residual covariance S and the Kalman gain K are evaluated and used to update the state and covariance matrix to obtain the *a posteriori* estimates. Note that in the case of the error state EKF, Q_d and H_k couple system noise and measurements with the error state $\Delta\mathbf{x}$ rather than $\hat{\mathbf{x}}$.

Although this EKF cycle can be repeated each time measurements arrive, for performance reasons we have chosen to group the incoming measurements to the highest frequency measurement, i.e., the IMU data. Hence, each time any non-IMU measurement arrives, it is slightly delayed until the next IMU measurement is available. The maximum possible sampling error caused by this grouping approach is $1/(2 \times 90)$ s and thus it can be neglected compared to the significantly longer sampling periods of the non-IMU data sources. The update rate of the EKF is then equal to the IMU sampling rate, i.e., 90 Hz.

4.3. Measurement Error Model

In general, the measurement vector $\mathbf{y}$ can be described as a sum of measurement function $h(\mathbf{x})$ of the state $\mathbf{x}$ and of some random noise $\mathbf{m}$ due to properties of the individual sensors:

$$\mathbf{y} = h(\mathbf{x}) + \mathbf{m}. \tag{12}$$

Using the function h, we can predict the measured value based on current knowledge about the system state:

$$\hat{\mathbf{y}} = h(\hat{\mathbf{x}}). \tag{13}$$

There is a difference $\Delta\mathbf{y} = \hat{\mathbf{y}} - \mathbf{y}$ caused by the modeling imperfections in the state estimate as well as by the sensor errors. This difference can be expressed in terms of the error state $\Delta\mathbf{x}$:

$$\begin{aligned}\Delta\mathbf{y} &= \mathbf{y} - \hat{\mathbf{y}} = h(\mathbf{x}) - h(\hat{\mathbf{x}}) + \mathbf{m} \\ &= h(\hat{\mathbf{x}} + \Delta\mathbf{x}) - h(\hat{\mathbf{x}}) + \mathbf{m}.\end{aligned} \tag{14}$$

If function h is linear, Eq. (14) becomes

$$\Delta\mathbf{y} = h(\Delta\mathbf{x}) + \mathbf{m}. \tag{15}$$

Although the condition of linearity is not always met, we still can approximate the behavior of h in some close proximity to the current state $\hat{\mathbf{x}}$ by a similar function h', which is linear in elements of $\hat{\mathbf{x}}$ such that

$$h(\hat{\mathbf{x}} + \Delta\mathbf{x}) - h(\hat{\mathbf{x}}) \approx h'(\Delta\mathbf{x})|_{\hat{\mathbf{x}}} = H_{\hat{\mathbf{x}}} \Delta\mathbf{x}, \tag{16}$$

where $H_{\hat{\mathbf{x}}}$ is the innovation matrix projecting observed differences in measurements onto the error states.

4.3.1. IMU Measurement Model

The inertial measurement unit is capable of measuring specific force (Savage, 1998) in all three dimensions as well as angular rates. The specific force measurement is a sum of acceleration and gravitational force, but it also contains biases—constant or slowly changing value independent of the actual acting forces—and sensor noise, which is

expected to have zero mean normal probability. All the values are measured in the I-frame,

$$\mathbf{y}_{f,I} = \mathbf{f}_I + \mathbf{b}_{f,I} + \mathbf{m}_{f,I}, \tag{17}$$

where $\mathbf{y}_{f,I}$ is the measurement, $\mathbf{f}_I$ is the true specific force, $\mathbf{b}_{f,I}$ is sensor bias, and $\mathbf{m}_{f,I}$ is sensor noise.

Since the interesting value $\mathbf{y}_{f,I}$ is expressed in the I-frame, we define a constant rotation matrix C_R^I of the R-frame to the I-frame. Translation between the I- and R-frames does not affect the measured values directly; thus, it is not considered. Since the IMU is placed close to the R-frame origin, we neglect centrifugal force induced by rotation of the R-frame and conditioned by nonzero translation between the R- and I-frames. Using this rotation matrix, we express the measurement as

$$\mathbf{y}_{f,I} = C_R^I \mathbf{f}_R + \mathbf{b}_{f,I} + \mathbf{m}_{f,I}, \tag{18}$$

where both $\mathbf{f}_R$ and $\mathbf{b}_{f,I}$ are elements of the system state. If we compare the measured value and the expected measurement, we can express the h function, which is—in this case—equal to the h':

$$\begin{aligned}\mathbf{y}_{f,I} - \hat{\mathbf{y}}_{f,I} &= \Delta\mathbf{y}_{f,I} = C_R^I \mathbf{f}_R + \mathbf{b}_{f,I} - C_R^I \hat{\mathbf{f}}_R - \hat{\mathbf{b}}_{f,I} + \mathbf{m}_{f,I} \\ &= C_R^I \Delta\mathbf{f}_R + \Delta\mathbf{b}_{f,I} + \mathbf{m}_{f,I},\end{aligned} \tag{19}$$

and hence can be expressed in $H_{\hat{\mathbf{x}}}\Delta\mathbf{x}$ form as

$$\Delta\mathbf{y}_{f,I} = \begin{bmatrix} \emptyset_3 & \emptyset_3 & \emptyset_3 & \emptyset_3 & C_R^I & \emptyset_3 & I \end{bmatrix} \Delta\mathbf{x} + \mathbf{m}_{f,I}, \tag{20}$$

where the error state $\Delta\mathbf{x}$ was defined in Eq. (1).

The angular rate measurement is treated identically; the output of the sensor is

$$\mathbf{y}_{\omega,I} = \boldsymbol{\omega}_I + \mathbf{b}_{\omega,I} + \mathbf{m}_{\omega,I}, \tag{21}$$

where $\boldsymbol{\omega}_I$ is the angular rate, $\mathbf{b}_{\omega,I}$ is sensor bias, and $\mathbf{m}_{\omega,I}$ is sensor noise.

Similarly, the measurement residual is obtained:

$$\mathbf{y}_{\omega,I} - \hat{\mathbf{y}}_{\omega,I} = \Delta\mathbf{y}_{\omega,I} = C_R^I \Delta\boldsymbol{\omega}_R + \Delta\mathbf{b}_{\omega,I} + \mathbf{m}_{\omega,I}, \tag{22}$$

which can be expressed in the matrix form

$$\Delta\mathbf{y}_{\omega,I} = \begin{bmatrix} \emptyset_3 & \emptyset_3 & \emptyset_3 & C_R^I & \emptyset_3 & \emptyset_3 & I \end{bmatrix} \Delta\mathbf{x} + \mathbf{m}_{\omega,I}. \tag{23}$$

4.3.2. Odometry Measurement Model

Our platform is equipped with caterpillar tracks and, therefore, steering is realized by setting different velocities to each of the tracks (*skid-steering*). The velocities are measured by incremental optical angle sensors at 15 Hz. Originally, we implemented a complex model introduced in Endo et al. (2007), which exploits angular rate measurements to model the slippage to further improve the odometry precision. However, with respect to our sensors, no improvement was observed. Moreover, since the slippage is inherently corrected via the proposed data fusion, we can neglect it in the odometry model, assuming only a very simple but sufficient model:

$$v_{O,x} = \frac{v_r + v_l}{2}, \tag{24}$$

where $v_{O,x}$ is the forward velocity, and v_l and v_r are track velocities measured by incremental optical sensors—the velocities in the lateral and vertical axes are set to zero. Since the robot position is obtained by integrating velocity expressed in the R-frame, we define a rotation matrix C_R^O:

$$\mathbf{v}_O = C_R^O \mathbf{v}_R, \tag{25}$$

which expresses the $\mathbf{v}_R$ in the O-frame.

During experimental evaluation, we observed a minor misalignment between these two frames, which can be described as rotation about the lateral axis by approximately one degree. Although relatively small, this rotation caused the position estimate in the vertical axis to grow at a constant rate while the robot was moving forward. To compensate for this effect, we handle the C_R^O as constant—its value was obtained by means of calibration. The measurement equation is then as follows:

$$\mathbf{y}_{v,O} = C_R^O \mathbf{v}_R + \mathbf{m}_{v,O}, \tag{26}$$

where $\mathbf{y}_{v,O}$ is linear velocity measured by the track odometry, expressed in the O-frame. Since this relation is linear, the measurement innovation is

$$\begin{aligned}\mathbf{y}_{v,O} - \hat{\mathbf{y}}_{v,O} &= \Delta\mathbf{y}_{v,O} \\ &= C_R^O \mathbf{v}_R - C_R^O \hat{\mathbf{v}}_R + \mathbf{m}_{v,O} \\ &= C_R^O \Delta\mathbf{v}_R + \mathbf{m}_{v,O}\end{aligned} \tag{27}$$

and expressed in the matrix form

$$\Delta\mathbf{y}_{v,O} = \begin{bmatrix} \emptyset_3 & \emptyset_3 & C_R^O & \emptyset_3 & \emptyset_3 & \emptyset_3 & \emptyset_3 \end{bmatrix} \Delta\mathbf{x} + \mathbf{m}_{v,O}. \tag{28}$$

4.3.3. ICP-based Localization Measurement Model

The ICP algorithm is used to estimate translation and rotation between each new incoming laser scan of the robot surroundings and a metric map created from the previously registered laser scans. In the course of our work, three approaches processing the output of the ICP were proposed and tested. The first approach treats the ICP-based localization as movement in the R-frame in between two consecutive laser scans in the form of a position increment (the *incremental position approach*). The idea of measurements expressed in a form of some $\Delta\mathbf{p}$ can be, for example, found in Ma et al. (2012). In our case, the increment is obtained as

$$\Delta\mathbf{p}_{R,\mathrm{ICP},i} = C_{(\mathbf{q}^R_{N,\mathrm{ICP},i-1})}(\mathbf{p}_{N,\mathrm{ICP},i} - \mathbf{p}_{N,\mathrm{ICP},i-1}), \tag{29}$$

where both the position $\mathbf{p}_{N,\mathrm{ICP}}$ and attitude $\mathbf{q}^R_{N,\mathrm{ICP}}$ are outputs of the ICP algorithm. The increment $\Delta\mathbf{p}_{R,\mathrm{ICP},i}$ is added to the position estimated by the whole fusion algorithm at

time-step $i-1$ to be used as a direct measurement of position. The same idea is applied in the case of attitude (an increment in attitude is extracted by means of quaternion algebra). The purpose is to overcome the ICP world frame drift. However, it is impossible to correctly discretize the system equations with respect to the laser scan sampling frequency ($\frac{1}{3}$ Hz). Also, the assumption of measurements being independent is violated by utilizing a previously estimated state to create a new measurement. Thus, corrections that propagate to the system state from this measurement tend to be inaccurate.

The second approach treats the ICP output as velocity in the R-frame (the *velocity approach*). We consider it a state-of-the-art practice utilized, for example, by Almeida and Santos (2013). The velocity is expressed in the N-frame first:

$$\mathbf{v}_{N,\mathrm{ICP}} = \frac{\mathbf{p}_{N,\mathrm{ICP},i} - \mathbf{p}_{N,\mathrm{ICP},i-1}}{t(i) - t(i-1)}, \tag{30}$$

where $t()$ is time corresponding to a time-step i. To express the velocity in the R-frame:

$$\mathbf{v}_{R,\mathrm{ICP}}(t) = C_{(\mathbf{q}^R_{R',\mathrm{ICP}}(t) \otimes \mathbf{q}^{R'}_{N,\mathrm{ICP},i-1})} \mathbf{v}_{N,\mathrm{ICP}}, \tag{31}$$

it is necessary to interpolate the attitude between $\mathbf{q}^R_{N,\mathrm{ICP},i-1}$ and $\mathbf{q}^R_{N,\mathrm{ICP},i}$ in order to obtain the increment $\mathbf{q}^R_{R',\mathrm{ICP}}(t)$. Angular velocity is assumed to be constant between the two laser scans. The velocity $\mathbf{v}_{R,\mathrm{ICP}}$ and the constant angular velocity obtained from the interpolation can be directly used as measurements that are independent of the estimated state, and because of the interpolation, they can be generated with arbitrary frequency and thus there is no problem with discretization (compared to the previous approach). However, this approach expects the robot to move in a line between the two ICP scans. This is a too strong assumption and also a major drawback of this approach, which results in incorrect trajectory estimates.

Therefore, we propose the third approach, the *trajectory approach*, which overcomes the assumption of the *velocity approach* by (suboptimal) use of the estimated states in order to approximate possible behavior of the system between each two consecutive ICP scans. This *trajectory approach* proved to be the best for preprocessing the output of the ICP algorithm; for details, see Section 5.4.5.

The *trajectory approach* assumes that the first estimate of the trajectory (without the ICP measurement) is locally very similar to the true trajectory (up to the effects of drift). Thus, when a new ICP measurement arrives, the trajectory estimated since the previous ICP measurement is stored to be used as the best guess around the previous ICP pose. The ICP poses at time-steps i and $i-1$ are aligned with the N-frame so the ICP pose at time-step $i-1$ coincides with the first pose of the stored trajectory. In this way, the ICP world frame drift is suppressed. Then, the stored trajectory is duplicated and aligned with the new ICP pose to serve as the best guess around the new ICP pose; see Figure 4. The resulting trajectory is obtained as the weighted average of the original and the duplicated trajectories:

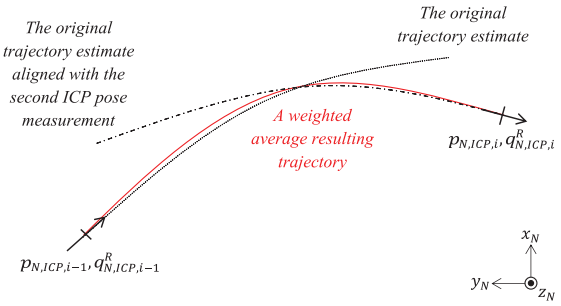


Figure 4. The principle of *trajectory approach*: when the new ICP measurement arrives (time-step i), the trajectory estimate based on measurements other than ICP (black dotted line) is duplicated and aligned with the incoming ICP measurement (black dashed line), and the weighted average (red solid line) of these two trajectories is computed.

$$\hat{\mathbf{p}}_{N,\mathrm{weighted},k} = \hat{\mathbf{p}}_{N,k} w_k + \hat{\mathbf{p}}'_{N,k} w'_k, \tag{32}$$

where $\hat{\mathbf{p}}_{N,k}$ are points of the original trajectory (black dotted line in Figure 4), $\hat{\mathbf{p}}'_{N,k}$ are points of the realigned duplicated trajectory (black dashed line in Figure 4), and w_k,w'_k are weights—linear functions of time equal to 1 at the time-step of associated ICP measurement and equal to 0 at the time-step of the other ICP measurement. The resulting trajectory is used to generate the velocity measurements in the N-frame as follows:

$$\mathbf{v}_{N,\mathrm{weighted},k} = \frac{\mathbf{p}_{N,\mathrm{weighted},k} - \mathbf{p}_{N,\mathrm{weighted},k-1}}{t(k) - t(k-1)}, \tag{33}$$

where $t(k)$ and $t(k-1)$ are the time-steps of poses of the resulting weighted trajectory. The k denotes indexing of the fusion algorithm high-frequency samples. Velocities can be expressed in the R-frame using the attitude estimates $\hat{\mathbf{q}}^R_{N,k}$:

$$\mathbf{v}_{R,\mathrm{weighted},k} = C_{(\hat{\mathbf{q}}^R_{N,k})} \mathbf{v}_{N,\mathrm{weighted},k}, \tag{34}$$

and they can be used directly as measurement, whose projection onto the error-state vector yields

$$\Delta\mathbf{y}_{v,\mathrm{weighted}} = \begin{bmatrix} \emptyset_3 & \emptyset_3 & I_3 & \emptyset_3 & \emptyset_3 & \emptyset_3 & \emptyset_3 \end{bmatrix} \Delta\mathbf{x} + \mathbf{m}_{v,\mathrm{weighted}}. \tag{35}$$

The velocity expressed in the R-frame can be used in this way as a measurement, but its values for the time period between two consecutive ICP outputs are known only *after* the second ICP measurement arrives. Thus it is necessary to recompute state estimates for this whole time period (typically in a length of 300 IMU samples), including the new velocity measurements.

To process the attitude information provided as the ICP output, we use a simple incremental approach such that the drift of the ICP world frame with respect to the N-frame is suppressed. To achieve this, we extract only the increment in attitude between two consecutive ICP poses:

$$\mathbf{q}^R_{N,\mathrm{ICP},i} = \mathbf{q}^R_{R',\mathrm{ICP}} \otimes \mathbf{q}^{R'}_{N,\mathrm{ICP},i-1}, \tag{36}$$

$$\mathbf{q}^R_{R',\mathrm{ICP}} = \mathbf{q}^R_{N,\mathrm{ICP},i} \otimes \left(\mathbf{q}^{R'}_{N,\mathrm{ICP},i-1}\right)^{-1}, \tag{37}$$

where $\mathbf{q}^R_{R',\mathrm{ICP}}$ is rotation that occurred between two consecutive ICP measurements, $\mathbf{q}^{R'}_{N,\mathrm{ICP},i-1}$ and $\mathbf{q}^R_{N,\mathrm{ICP},i}$. We apply this rotation to the attitude state estimated at time-step $k' \equiv i-1$:

$$\mathbf{y}_{q,\mathrm{ICP}} = \mathbf{q}^R_{R',\mathrm{ICP}} \otimes \hat{\mathbf{q}}^R_{N,k'}. \tag{38}$$

To express the measurement residual, we define the following error quaternion:

$$\delta\mathbf{q}_{\mathrm{ICP},i} = \hat{\mathbf{q}}^R_{N,k} \otimes \left(\mathbf{y}_{q,\mathrm{ICP}}\right)^{-1}, \tag{39}$$

where $\hat{\mathbf{q}}^R_{N,k}$ is the attitude estimated at time-step $k \equiv i$. We express this residual rotation by means of rotation vector $\delta\boldsymbol{\theta}_{\mathrm{ICP},i}$,

$$\delta\boldsymbol{\theta}_{\mathrm{ICP},i} = 2\vec{\delta\mathbf{q}}_{\mathrm{ICP},i}, \tag{40}$$

which can be projected onto the error state as

$$\Delta\mathbf{y}_{\delta\theta,\mathrm{ICP}} = \begin{bmatrix} \emptyset_3 & I_3 & \emptyset_3 & \emptyset_3 & \emptyset_3 & \emptyset_3 & \emptyset_3 \end{bmatrix} \Delta\mathbf{x} + \mathbf{m}_{\delta\theta,\mathrm{ICP}}. \tag{41}$$

Although the ICP is very accurate in measuring translation between consecutive measurements, the attitude measurement is not as precise. Noise introduced in the pitch angle can cause wrong velocity estimates expressed in the R-frame, resulting in a problem described as *climbing robot*—the system tends to slowly drift in the vertical axis. Since the output of the *trajectory approach* is velocity $\mathbf{v}_{R,\mathrm{weighted},i}$, applying a constraint assuming only planar motion in the R-frame is fully justified, easy to implement, and resolves this issue.

4.3.4. Visual Odometry Measurement Model

As explained in Section 3.5, the VO is an algorithm for estimating translation and rotation of a camera body based on images recorded by the camera. The current implementation of the data fusion utilizes only the rotation part of the motion estimated by the VO, since it is not affected by the scale. The set of 3D landmarks maintained by the VO is not in any way processed by the fusion algorithm—it is used by the VO to improve its attitude estimates internally. Similarly, the bundle adjustment ensures more consistent measurements, yet still, it does not enter the data fusion models.[3] The way we incorporate the VO measurements is equivalent to the ICP *trajectory approach*, however, reduced only to the incremental processing of the attitude measurements. In this way, the whole VO processing block can easily be replaced by an alternative (for example, by stereovision-based VO), provided the output—the estimated rotation—is available in the same way. The motivation is to have the VO measurement model independent of the VO internal implementation details. The implementation of the VO attitude aiding is identical to the ICP attitude aiding; the attitude increment is extracted and used to construct a new measurement $\mathbf{y}_{q,\mathrm{VO}}$:

$$\mathbf{q}^R_{N,\mathrm{VO},i} = \mathbf{q}^R_{R',\mathrm{VO}} \otimes \mathbf{q}^{R'}_{N,\mathrm{VO},i-1}, \tag{42}$$

$$\mathbf{q}^R_{R',\mathrm{VO}} = \mathbf{q}^R_{N,\mathrm{VO},i} \otimes \left(\mathbf{q}^{R'}_{N,\mathrm{VO},i-1}\right)^{-1}, \tag{43}$$

where $\mathbf{q}^R_{R',\mathrm{VO}}$ is rotation that occurred between two consecutive VO measurements $\mathbf{q}^{R'}_{N,\mathrm{VO},i-1}$ and $\mathbf{q}^R_{N,\mathrm{VO},i}$. We apply this rotation to the attitude state estimated at time-step $k' \equiv i-1$:

$$\mathbf{y}_{q,\mathrm{VO}} = \mathbf{q}^R_{R',\mathrm{VO}} \otimes \hat{\mathbf{q}}^R_{N,k'}. \tag{44}$$

Then, the measurement residual is expressed as an error quaternion:

$$\delta\mathbf{q}_{\mathrm{VO},i} = \hat{\mathbf{q}}^R_{N,k} \otimes \left(\mathbf{y}_{q,\mathrm{VO}}\right)^{-1}, \tag{45}$$

where $\hat{\mathbf{q}}^R_{N,k}$ is the attitude estimated at time-step $k \equiv i$. We express this residual rotation by means of rotation vector $\delta\boldsymbol{\theta}_{\mathrm{VO},i}$,

$$\delta\boldsymbol{\theta}_{\mathrm{VO},i} = 2\vec{\delta\mathbf{q}}_{\mathrm{VO},i}, \tag{46}$$

which can be projected onto the error state as

$$\Delta\mathbf{y}_{\delta\theta,\mathrm{VO}} = \begin{bmatrix} \emptyset_3 & I_3 & \emptyset_3 & \emptyset_3 & \emptyset_3 & \emptyset_3 & \emptyset_3 \end{bmatrix} \Delta\mathbf{x} + \mathbf{m}_{\delta\theta,\mathrm{VO}}, \tag{47}$$

where $\mathbf{m}_{\delta\theta,\mathrm{VO}}$ is the VO attitude measurement noise.

5. EXPERIMENTAL EVALUATION

Our evaluation procedure involves several different tests. First, we describe our evaluation methodology in Section 5.1. It covers obtaining ground-truth positioning measurements for both indoors and outdoors. Then we present and discuss our field experiments with the global behavior of our state estimation (Section 5.2). We also show two examples of typical behavior of the filter in order to give more insight on its general characteristics (Section 5.3). We

[3]The same idea applies for the ICP-based localization: although it builds an internal map, this map is independent from our localization estimates. This would not be the case in a SLAM approach with integrated loop closures.

Figure 5. The experimental setup with the Leica reference theodolite for obtaining ground truth trajectory (left). Part of the 3D semistructured environment for an indoor test with motion capture ground truth (right).

take advantage of them to explain the importance of the *trajectory approach* compared to more standard measurement models. Finally, we analyze the behavior of the filter under failure case scenarios involving partial or full outage of each sensory modality (Section 5.4).

5.1. Evaluation Metrics

To validate the results of our fusion system, we need accurate measurements of part of our system states to confront with the proposed filter. For indoor measurements, we use a Vicon motion capture system with nine cameras covering more than 20 m^2 and giving a few millimeter accuracy at 100 Hz.

For external tracking, we use a theodolite from Leica Geosystems, namely the Total Station TS15; see Figure 5 (left). It can track a reflective prism to measure its position continuously at an average frequency of 7.5 Hz. The position precision of the theodolite is 3 mm in continuous mode. However, this system cannot measure the orientation of the robot. Moreover, the position measured is that of the prism and not directly of the robot, therefore we calibrated the position of the prism with respect to the robot body using the theodolite and precise blueprints. However, the position of the robot cannot be recovered from the position of the prism without the information about orientation. That explains why, in the validations below, we do not compare the position of the robot but rather the position of the prism from the theodolite and reconstructed from the states of our filter. With these ground-truth measurements, we use different metrics for evaluation. First, we simply plot the error as a function of time. More precisely, we consider *position error*, *velocity error*, and *attitude error* and we compute them by taking the norm of the difference between the prediction made by our filter and the reference value.

Since this metric shows how the errors evolve over time, a more condensed measure is needed to summarize and compare the results of different versions of the filter. Therefore, we use the *final position error* expressed as a percentage of the total trajectory length:

$$e_{\text{rel}} = \frac{||\mathbf{p}_l - \mathbf{p}_{\text{ref},l}||}{\text{distance traveled}}, \tag{48}$$

where l is the index of the last position sample $\mathbf{p}_l$ with the corresponding reference position $\mathbf{p}_{\text{ref},l}$.

While this metric is convenient and widely used in the literature, it is, however, representative only of the end point error regardless of the intermediary results. This can be misleading for long trajectories in a confined environment as the end point might be close to the ground truth by chance. This is why we introduce, as a complement, the *average position error*:

$$e_{\text{avg}}(l) = \frac{\sum_{i=1}^{l} ||\mathbf{p}_i - \mathbf{p}_{\text{ref},i}||}{l}, \tag{49}$$

where $1 \leq l \leq$ *total number of samples*. To improve the legibility of this metric in the plots, we express the e_{avg} as a function of time,

$$e'_{\text{avg}}(t) = e_{\text{avg}}(l(t)), \tag{50}$$

where $l(t)$ simply maps time t to the corresponding sample l.

5.2. Performance Overview of the Proposed Data Fusion

With these metrics, we can actually evaluate the performance of our system in a quantitative way. We divided the tests into indoor and outdoor experiments.

5.2.1. Indoor Performance

For the indoor tests, we replicated a semistructured environment found in USAR environments, including ramps, boxes, a catwalk, a small passage, etc. Figure 5 (right) shows a picture of part of the environment. Due to the limitations of our motion capture setup, this testing environment is

Table II. Comparison of combinations of different modalities evaluated on indoor experiments performed under standard conditions with the Vicon system providing ground truth in position and attitude. Final position error expressed in percent of the total distances traveled was chosen as a metric for each experiment; the total distance of the 28 experiments was 765 m, including traversing obstacles.

			Final position error in % of the distance traveled			
Exp.	Distance traveled (m)	Exp. duration (s)	OD, IMU	OD, IMU, VO	OD, IMU, ICP	OD, IMU, ICP, VO
1	47.42	254	2.17	2.30	1.71	0.79
2	36.52	186	1.99	2.21	0.36	0.14
3	48.74	244	3.15	2.63	0.50	0.18
4	29.40	237	2.22	2.06	0.42	0.45
5	82.10	585	2.51	2.24	0.90	0.71
6	74.64	452	2.05	3.64	0.98	1.24
7	74.65	387	1.70	1.72	2.28	0.58
8	30.57	194	1.98	3.42	1.59	2.29
9	26.58	287	2.67	2.23	1.90	1.19
10	26.57	236	1.53	3.94	0.77	2.11
11	26.96	208	1.25	1.20	0.95	0.66
12	29.13	211	1.27	1.29	0.88	0.87
13	26.35	180	1.37	1.25	0.94	0.77
14	40.23	240	6.58	6.70	0.88	0.99
15	21.01	167	5.26	5.27	0.61	0.57
16	19.04	209	5.94	5.95	0.55	0.60
17	10.95	405	3.44	2.89	2.15	2.05
18	8.65	238	2.87	2.77	1.36	1.38
19	9.36	284	4.14	3.91	1.83	1.85
20	9.02	282	2.90	3.36	2.73	2.65
21	10.82	308	3.79	3.23	1.43	1.41
22	9.45	237	5.36	5.45	2.66	2.68
23	12.75	204	2.65	2.84	2.66	1.79
24	7.81	179	1.58	1.83	2.82	3.06
25	10.85	165	3.85	4.14	3.25	2.17
26	10.83	163	2.36	1.84	0.62	0.68
27	12.79	237	15.42	14.95	2.48	2.53
28	12.07	239	28.42	27.07	2.89	2.98
Lower quartile | **Median** | Upper quartile			2.0|**2.7**|4.0	2.1|**2.9**|4.0	0.8|**1.4**|2.4	0.7|**1.2**|2.1

not as large as typical indoor USAR environments. Nevertheless, it features most of the complex characteristics that make state estimation challenging in such an environment.

For this evaluation, we recorded approximately 2.4 km of indoor data with ground truth; 28 runs represent standard conditions (765 m in total), and 36 runs represent failure cases of different sensory modalities induced artificially (1,613 m in total). Table II presents the results of each combination of sensory modalities for the 28 standard conditions runs; the failure scenarios are analyzed in Section 5.4 separately.

The sensory modality combinations can be divided into two groups by including or excluding the ICP modality; these two groups differ by the magnitude of the final position error. From this fact, we conclude that the main source of error is slippage of the caterpillar tracks—the VO modality in our fusion system corrects only the attitude of the robot. Also, the results confirmed sensitivity to erroneous attitude measurements originating from the sensory modalities. In this instance, VO slightly worsened the median of the final position error—the indoor experiments are not long enough to make the difference between drift rates of the bare IMU+OD combination and possible VO errors that originate from incorrect pairing of image features. Nevertheless, the results are not significantly different.[4] A significant improvement is achieved with the ICP modality, which compensates for the track slippage and reduces the resulting median of the final position errors by 50% (approximately). As expected during the

[4] All statistically significant results are assessed using the Wilcoxon signed-rank test with $p < 0.05$ testing whether the median of correlated samples is different.

Figure 6. Pictures of the outdoor environments in Zurich. Left: street canyon, right: urban park.

Table III. Comparison of combinations of different modalities evaluated on outdoor experiments performed under standard conditions with the Leica system providing ground truth in position.

			Final position error in % of the distance traveled			
Experiment	Distance traveled (m)	Exp. duration (s)	OD, IMU	OD, IMU, VO	OD, IMU, ICP	OD, IMU, ICP, VO
1: basement 1	120.62	825	2.08	26.61	1.83	17.84
2: basement 2	175.67	853	1.37	12.53	2.42	5.91
3: hallway straight	159.42	738	1.10	20.48	0.43	12.22
4: street 1	135.18	584	2.78	0.72	0.24	0.62
5: street 2	259.86	992	9.74	0.80	0.26	0.80
6: park big loop	145.31	918	2.65	2.66	1.03	1.76
7: park small loop	88.20	601	1.94	1.60	1.25	0.97
8: park straight	99.29	560	1.20	20.18	0.62	11.50
9: 2 floors	238.28	1010	9.10	0.62	0.58	0.43
10: 2 floors opposite	203.23	1107	3.23	6.79	0.51	0.42
Lower quartile | **Median** | Upper quartile			1.4|**2.4**|3.2	0.8|**4.7**|20.2	0.4|**0.6**|1.2	0.6|**1.4**|11.5

filter design, fusing all sensory modalities yields the best result (not significantly different from that without VO), with a median of 1.2% final position error; the occasional VO attitude measurement errors are diminished by the ICP modality attitude measurement (and vice versa).

5.2.2. Outdoor Performance

We ran outdoor tests in various environments, namely a street canyon and an urban park with trees and stairs in Zurich. Figure 6 shows pictures of the environments.

In those environments, we recorded in total approximately 2 km, with ground truth available for 1.6 km; the rest were returns from the experimental areas. These 1.6 km are split into 10 runs. Table III, likewise Table II, presents the results of each combination of sensory modalities for each run.

Contrary to the indoor experiments, combining all four modalities does not improve the precision of localization compared to ICP, IMU, and odometry fusion (the fusion of all is significantly worse than ICP, IMU, and odometry only). Although some runs show improvement while combining all the sensory modalities (runs 7, 9, and 10) or are at least comparable with the best result 0.4|**0.6**|1.2 (runs 4, 5, and 6), there were several experiments in which VO failed due to the specificities of the environments. Such failures result in erroneous attitude estimates significantly exceeding expected VO measurement noise and compromising the localization accuracy of the fusion algorithm. The reasons for the failures are described in the Section 5.4 together with other failure cases. Since we did not artificially induce these VO failures, as we did in the case of the indoor experiments, we do not exclude these runs from the performance evaluation in Table III—we consider such environments standard for USAR. Moreover, we treat them as additional proof of the fusion algorithm sensitivity to erroneous attitude measurement originating both from VO and ICP modalities, and we will address them in the conclusions and future work.

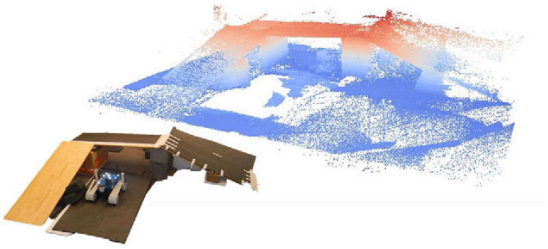


Figure 7. The 3D structure for testing obstacle traversability shown as a metric map created by ICP.

5.3. In-depth Analysis of the Examples of Performance

To provide more insight into the characteristics of the filter, we selected some trajectories, and we present more information than just the final position error metric.

5.3.1. Example of Data Fusion Performance in an Indoor Environment

In this example, we address the caterpillar track slippage when traversing an obstacle (Figure 7). Since we are looking forward to USAR missions, such environment with conditions inducing high slippage can be expected, e.g., collapsed buildings full of debris and dust that impair traction on smooth surfaces such as exposed concrete walls or floors, mass traffic accidents with oil spills, etc. The Vicon system was used to obtain precise position and orientation ground truth for computing the *average position error* development in time.

When traversing a slippery surface, any track odometry inevitably fails with the tracks moving with significantly diminishing traction. For this reason, trajectory and state estimates resulting from the IMU+OD fusion showed unacceptable error growth; see Figure 8. The robot was operated to attempt to climb up the yellow slippery board (Figure 7), which deteriorated the traction to the point that the robot was sliding back down with each attempt to steer. Because of the slippage, it failed to reach the top. Then, it was driven around the structure and up, to further slowly slip down the slope backward, with the tracks moving forward to spoil traction. The effect of the slippage on the OD is apparent from the purple line in Figure 8. The corresponding average position error of the bare combination of IMU+OD starts to build up as soon as the robot enters the slippery slope. At 75 s, the IMU+OD has already an error of 0.5 m and finishes at 200 s at an error of 4.4 m (outside Figure 8). Without exteroceptive modalities this problem is unsolvable, and, as expected, including these modalities significantly improves the localization accuracy; the final average position error is only 0.14 m for the IMU+OD+VO+ICP combination. The resulting state estimates for the combination of all modalities are shown in Figures 9 and 10. Figure 9 depicts position estimates (the upper left quarter) with the reference values. The difference between the estimate and the reference is plotted in the bottom left quarter; similarly, the right half of the figure displays the velocity estimate. In the left part of Figure 10, the attitude estimate expressed in Euler angles is shown with its error compared to the Vicon reference. The right part of this figure demonstrates estimation of the sensor biases, which are part of the system state. Note that the biases in angular rates are initialized to values obtained as the mean of angular rate samples measured when the robot remains stationary before each experiment—short self-calibration. In conclusion, adding the exteroceptive sensor modalities—as proposed in our filter design—compensated for the effect caused by high slippage shown in this example, as shown by the shape of trajectories and the average position error.

5.3.2. Example of Data Fusion Performance in an Outdoor Environment

This outdoor experiment took place on the Clausiusstrasse street (near ETH in Zurich) (Figure 11), and the purpose was to test the exteroceptive modalities (the ICP and the VO) in an open urban space. In this standard setting, both the ICP and the VO are expected to perform reasonably well, although the ICP—compared to a closed room—is missing a significant amount of spatial information (laser range is limited to approximately 50 m, no ceiling, etc.). The Leica theodolite was used to obtain the ground-truth position during this experiment (Figure 5).

The results are shown in Figures 12 and 13, and they demonstrate the improvement of performance when including more modalities up to the full setup. The basic dead-reckoning combination (IMU+OD) showed a clear drift in the yaw angle caused by accumulating error due to angular rate sensor noise integration (see the purple trajectory in the left part of Figure 12). By including the VO attitude measurements (resulting in IMU+OD+VO), the drift was compensated. Although the VO is not in fact completely drift-free, the performance is clearly better than the angular rate integration—rather it is the scale of the trajectory that matters. The IMU+OD+VO modality combination suffered from inaccurate track odometry velocity measurements (the green line in Figure 12), but this problem was resolved by incorporating the ICP modality into the fusion scheme. The IMU+OD+ICP+VO combination proved to provide the best results; see the average position error plot in Figure 12 (right). The attitude estimates and estimates of the sensor biases are shown in Figure 14.

5.3.3. Evaluation of the Measurement Model

We claim that a standard measurement model—as is usually used for measurements coming at comparable

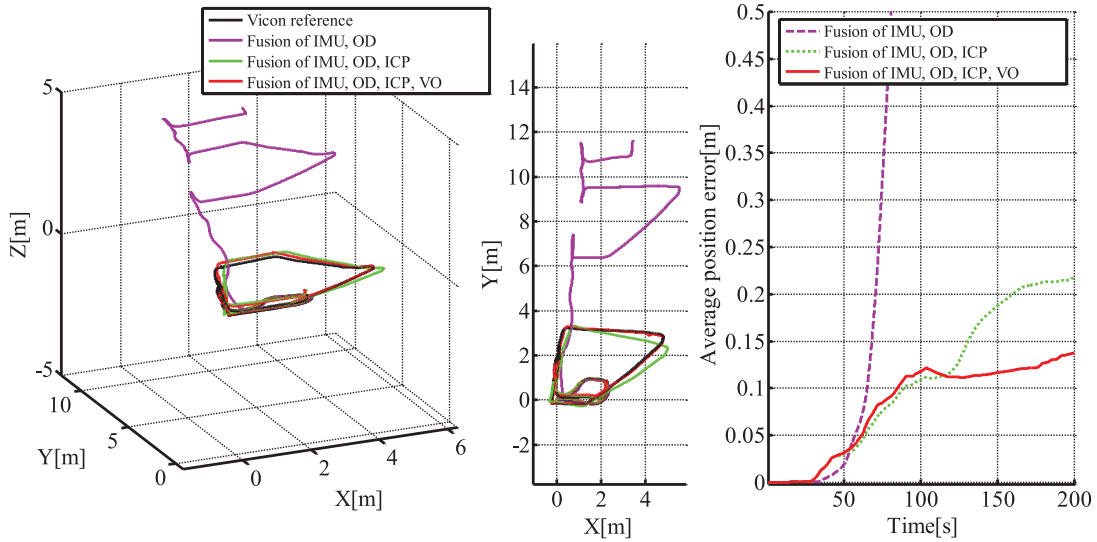


Figure 8. Trajectories obtained by fusing different combinations of modalities during the indoor experiment testing obstacle (depicted in Figure 7) traversability under high slippage (left, middle); development of the average position error (right).

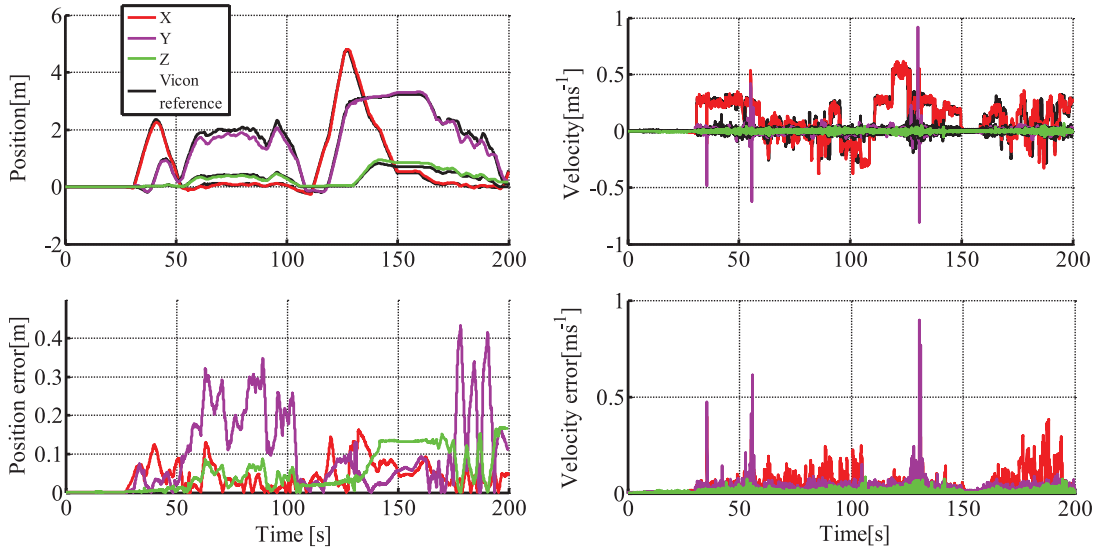


Figure 9. The corrected position (top left) and velocity estimates (top right) for the IMU+OD+ICP+VO combination corresponding to the trajectory in Figure 8 (testing obstacle traversability). Errors in position and velocity are obtained as the norm of difference between the Vicon reference and the corresponding state at each time-step (bottom left, bottom right). The Vicon reference for both position and velocity is shown in black.

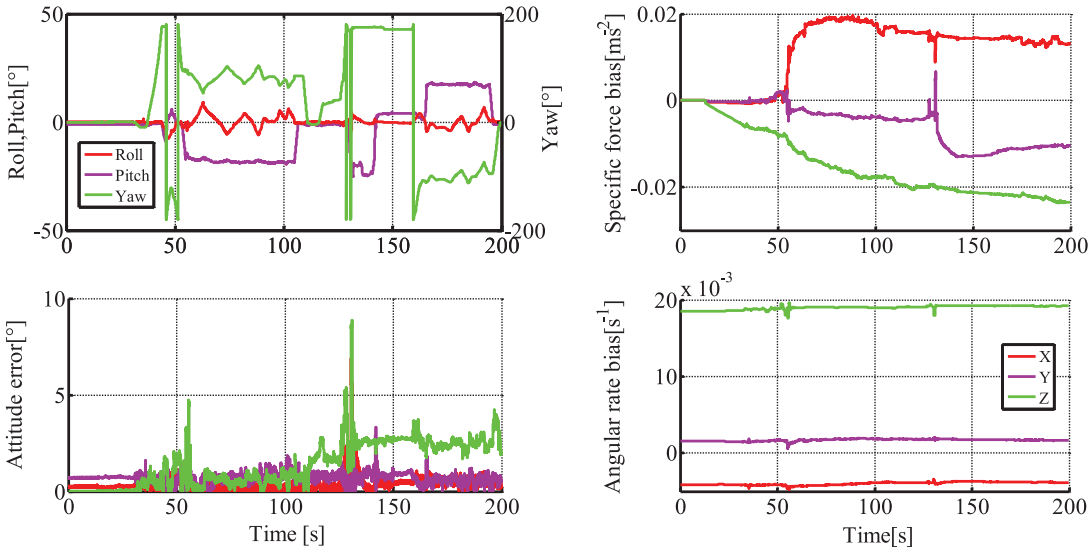


Figure 10. The corrected attitude estimates (top left) for the full multimodal combination IMU+OD+ICP+VO corresponding to the trajectory shown in Figure 8 (testing obstacle traversability). Errors in attitude are obtained as the difference between the Vicon reference and the corresponding state at each time-step (bottom left). Estimated biases for the specific forces (top right) and angular rates (bottom right).

Figure 11. An example of trajectory driven by the robot over the Clausiusstrasse street.

frequency—is not well-suited for measurements with significant differences in sampling frequencies as well as in values that correspond to the same state observed. This is crucial when the difference in states obtained from the IMU or the OD at high frequency is very large compared to the measurements provided by the ICP or the VO sensory modalities at relatively low frequency—such as in the case of high slippage.

Table IV shows the overall comparison of the three measurement models we evaluated for fusing the ICP and the VO sensory modalities in the filter. Figure 15 presents a typical example of trajectory reconstructed by all three measurement approaches we introduced in Section 4.3.3. The *velocity approach*—the state-of-the-art practice—that considers that information as relative measurements, is the least precise, with the highest average position error; see Figure 15 (right). This is due to the *corner cutting* behavior emphasized in Figure 15 (middle). The *incremental position approach* performs reasonably well in indoor environments, which are well-conditioned for the ICP and the VO sensory modalities. In particular, the ICP algorithm is very precise as there are enough features to unambiguously fix all degrees of freedom. On the other hand, in larger environments with fewer constraints (expected for USAR), the *trajectory approach* allows the IMU and the OD information to better correct the drift of the ICP and the VO sensory modalities.

5.4. Failure Case Analysis

As seen in the previous sections, there are many occasions in USAR environments for which the generic assumptions of the EKF are not valid. The most frequent example is track slippage, which violates the assumption of Gaussian observation centered on the actual value.

Our failure case analysis reviews each sensory modality involved in the filter to see how the resulting estimate degrades with partial outage of the modality. IMUs are not subject to much partial failure other than bias and noise, which are already accounted for in our filter.

5.4.1. Robot Slippage and Sliding

A typical failure case of the odometry modality is significant slippage. Small slippage occurs routinely when turning

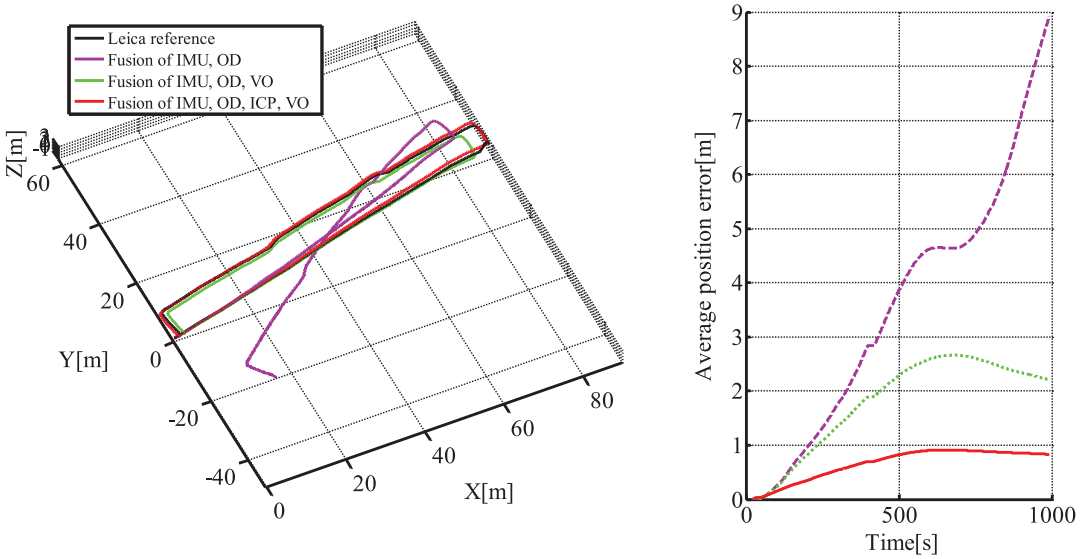


Figure 12. Trajectories obtained by fusing different combinations of modalities during the outdoor experiment with Leica reference system (left) and the corresponding average position error in time (right).

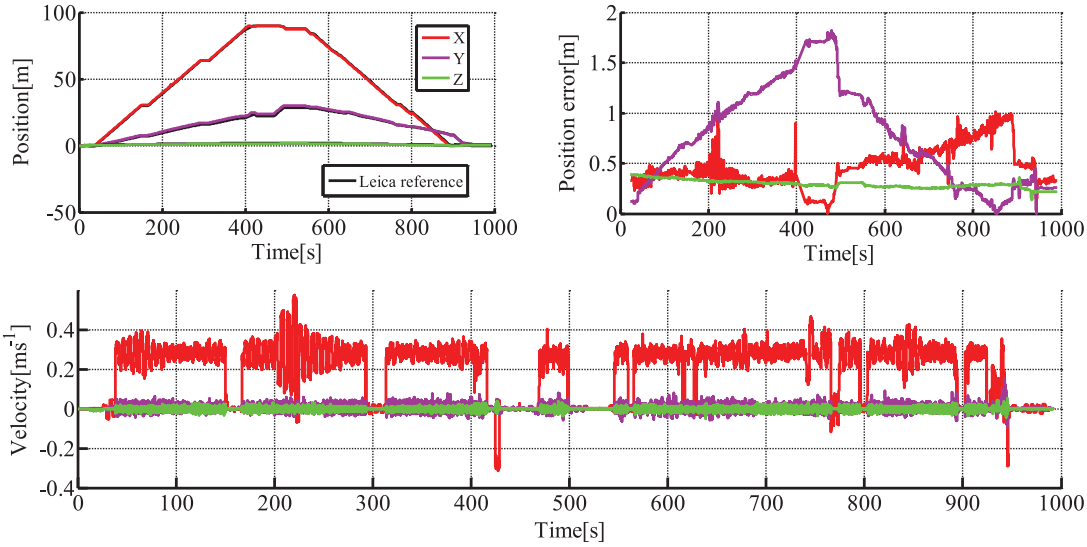


Figure 13. The position and velocity estimates (top left and bottom, respectively) for the IMU+OD+ICP+VO combination corresponding to the outdoor trajectory in Figure 12; errors in position obtained as the norm of differences between the Leica reference and the corresponding state at each time-step (top right).

skid-steer robots and is usually accounted for by the uncertainty in the odometry model. However, on surfaces such as ice, or inclined wet or smooth surfaces, stronger slippage can occur. Stronger slippage or sliding are outliers of the odometry observation model. IMU, ICP, and VO sensory modalities are not affected in such a case. To simulate such a situation, we placed the robot on a trolley and moved it manually.

Figure 16 shows both the trajectory from the top (top-left plot) and the comparison between the fusion of all four

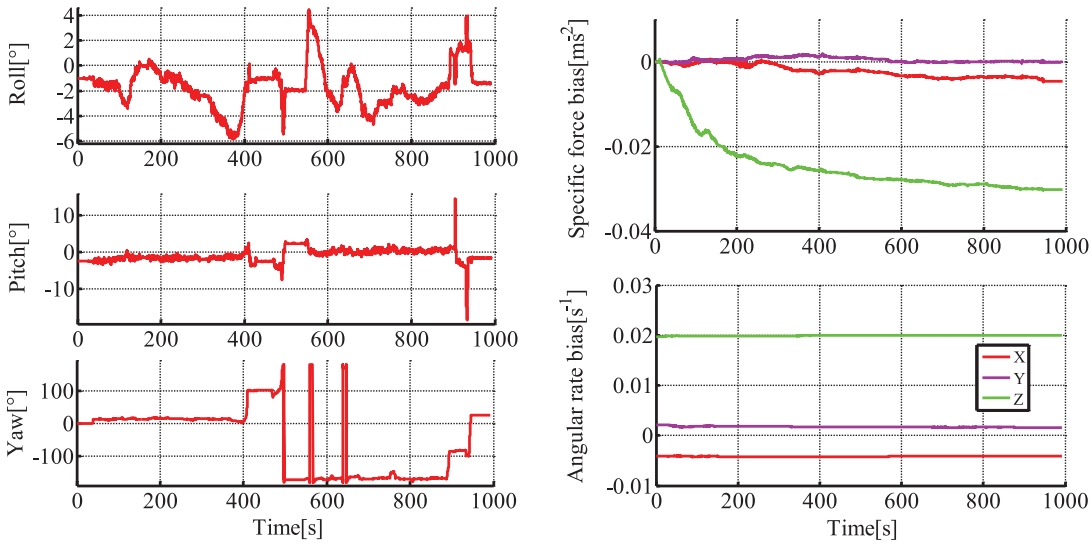


Figure 14. The attitude estimates (left) for the IMU+OD+ICP+VO combination corresponding to the outdoor trajectory in Figure 12; biases estimated for the specific forces (top right) and angular rates (bottom right).

Table IV. Comparison of the different measurement models; for each model, we show the lower|**median**|higher quartile statistics of the relative and average metrics. The average metric e_{avg} is evaluated for the last sample of each experiment; see Eq. (49). We distinguish the indoor and outdoor environments.

	Indoor		Outdoor	
Model	e_{rel}	e_{avg}	e_{rel}	e_{avg}
incremental position	0.4|**0.7**|1.2	0.1|**0.1**|0.2	0.8|**1.5**|11.0	0.7|**2.4**|6.1
velocity	1.0|**1.3**|2.3	0.1|**0.1**|0.3	0.9|**1.8**|12.2	0.8|**2.5**|6.1
trajectory	0.7|**1.2**|2.1	0.0|**0.1**|0.2	0.6|**1.4**|11.5	0.6|**2.2**|6.1

sensory modalities and the fusion of only IMU+OD. We can see that the latter wrongly estimates no motion, whereas the fusion of all modalities correctly estimates the trajectory. The failure of the partial filter can be explained by the low acceleration of the platform during the test. As the IMU acceleration signal is quite noisy, confidence in the IMU cannot compensate for the odometry modality asserting an absence of motion.

It should be noted that such a failure of the odometry modality does not lead to a failure of our complete filter.

5.4.2. Partial Occlusion of the Visual Field of View

Partial occlusion, overexposure, or projections of dirt on the camera could lead to faulty estimation of the motion by the VO. To test this situation, we occluded one of the cameras of the omnicamera (see Figure 17). Reduction of the field of view of the omnicamera causes in the vast majority of cases a reduction in the number of visual features being robustly detected by the VO. The insufficient number of features can then cause the VO to incorrectly estimate the attitude. This information then propagates into the state estimate and can cause the fusion algorithm to fail.

Figure 18 shows the result of the filter in such a case. We can see that during a first loop of the trajectory, the state estimation is correct. Then, lacking a sufficient number of features, the VO computes an erroneous estimate and the final state estimate degenerates. On the contrary, by leaving out the visual odometry, the state estimation would continue to perform satisfactorily.

It should be noted that the number, quality, and distribution of features matter more than the portion of the field of view that is occluded. One typical way to prevent this issue is to monitor the number of features and eventually their distribution in the field of view—our VO tries to have corresponding features spread over the whole image.

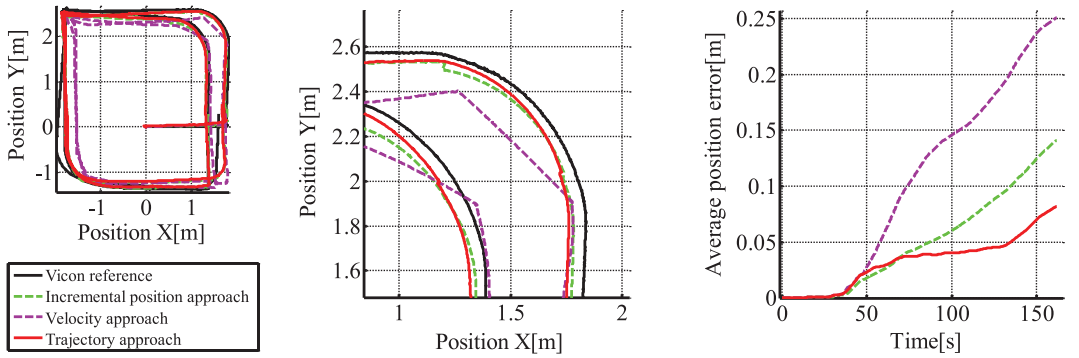


Figure 15. Comparison of effects of the three different ICP aiding approaches on the estimated trajectory (left, middle) and on the average position error (right). Note the *corner cutting* effect of the *velocity approach*.

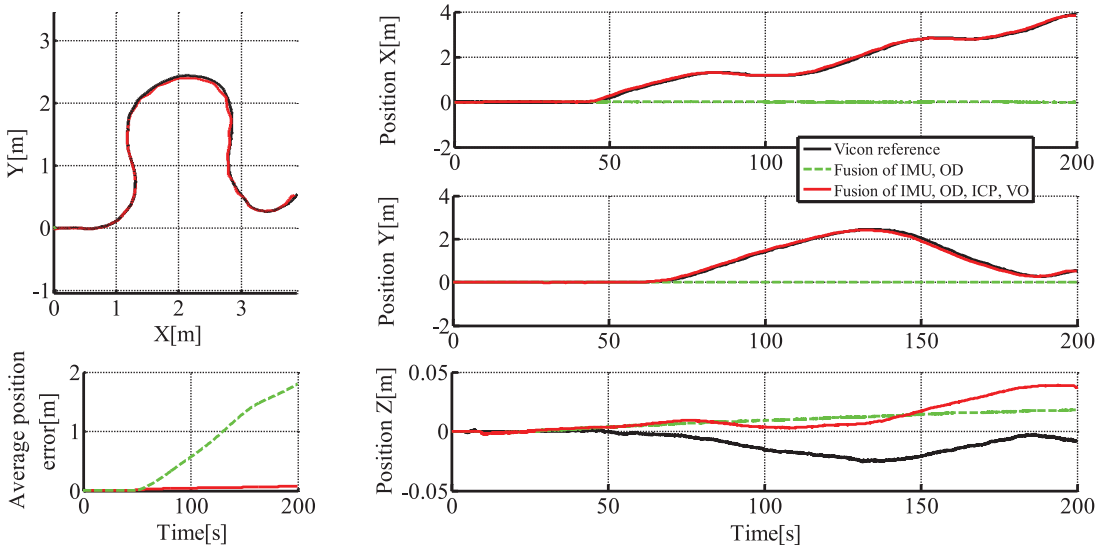


Figure 16. Test trajectory for robot slippage. Black line: ground truth; red solid line: state estimate with all four modalities; green dashed line: IMU and odometry fusion. Top left: top view of the trajectory; bottom left: average error as a function of time; top, middle, bottom right: evolution of x, y, and z coordinates.

5.4.3. Temporary Laser Scanner Outage

As demonstrated above, our trajectory approach to fusion of ICP measurements is able to cope with the relatively low frequency of laser scanning. As the laser is moving, it can be blocked in the case of collision or high vibration of the platform (a safety precaution at the level of the motor controller). When this happens, it is necessary to initiate a recalibration procedure that can take around 30 s.

We simulated this situation by throttling the laser point clouds, which resulted in ICP measurement outages of up to 40 s. Figure 19 shows the trajectory estimates for this test. On the left, the cyan polygon shows the position estimates of ICP linked by straight lines (no filtering). It should be noted that in this case, the positions are accurate compared to the ground truth but of very low sampling rate. We can see in the middle and right graphs that the filter estimates degrade gracefully. There is some drift, mostly along elevation due to slippage, but even with this low frequency, the ICP measurements help to correct the state estimates over just the IMU, odometry, and visual odometry.

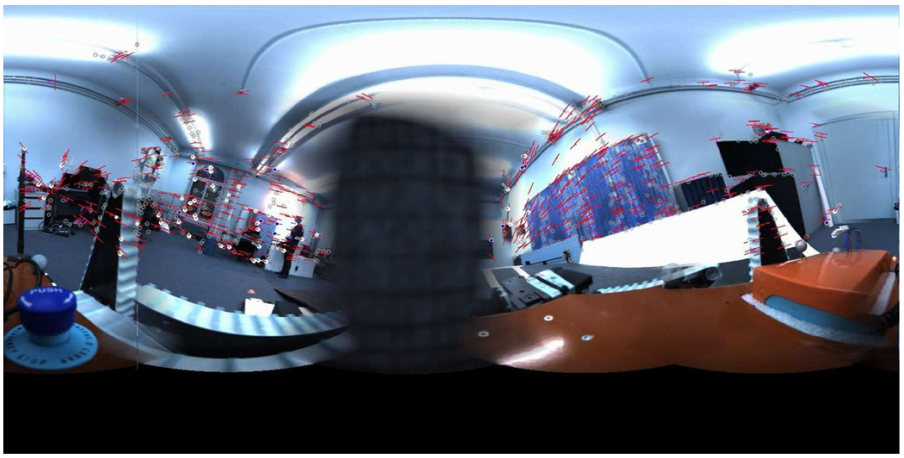


Figure 17. Picture from the partially occluded omnicamera. Notice the dark rectangle in the middle.

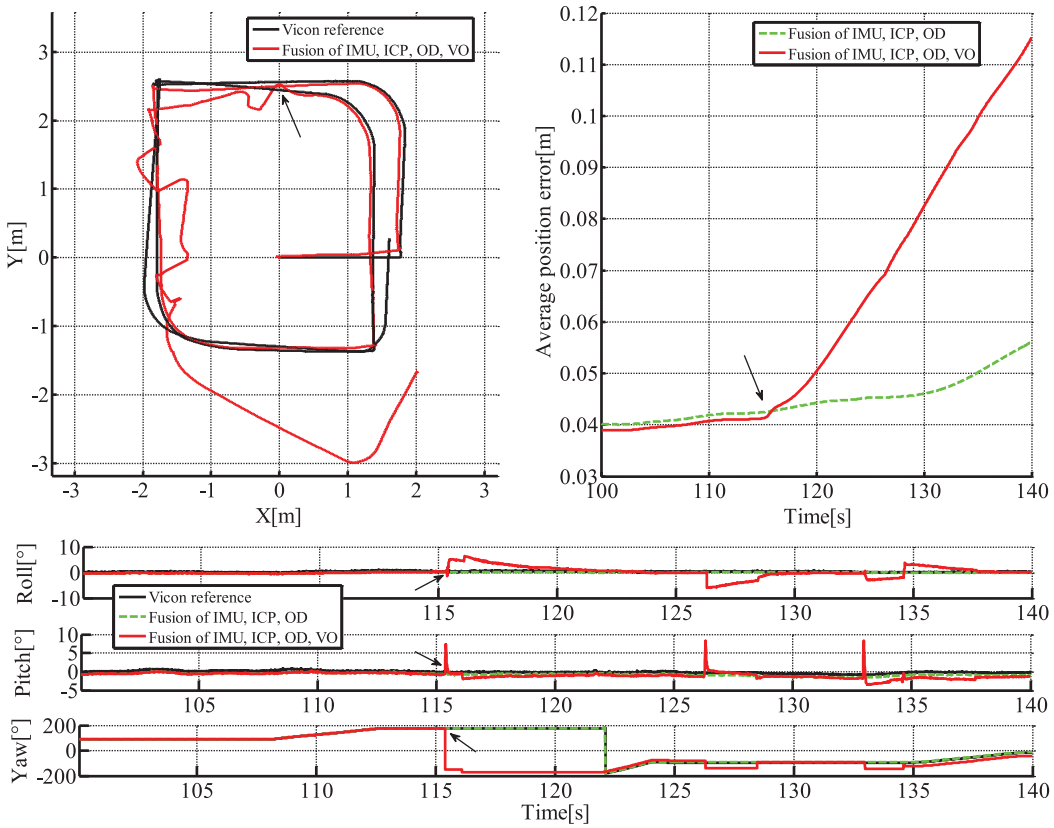


Figure 18. Trajectory reconstruction with several faulty VO motion estimates. Black line: ground truth; solid red line: state estimate with all four modalities; dashed green line: state estimate excluding visual odometry; black arrow: visual odometry failure. Top left: top view of the trajectory; top right: average position error around visual odometry failure; bottom: attitude estimated along the trajectory.

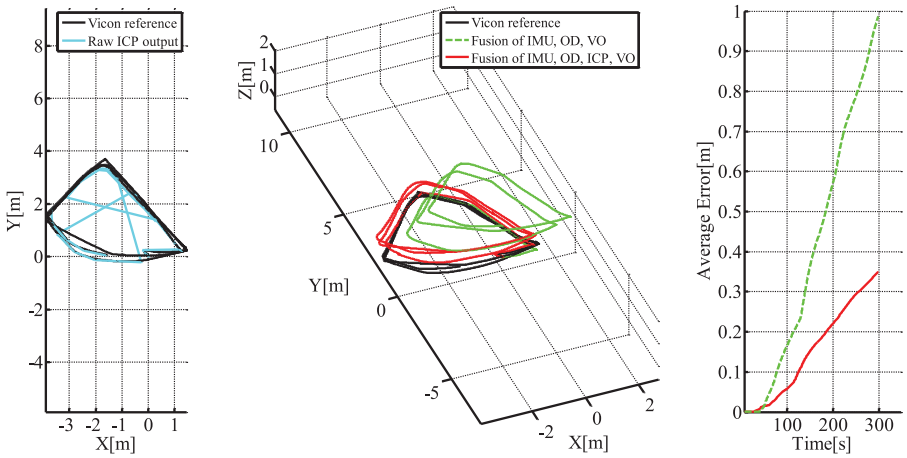


Figure 19. Trajectory estimates in the case of low ICP frequency. Black line: ground truth; cyan line: positions estimated by ICP alone; red line: state estimate with all four modalities; green dashed line: state estimate excluding ICP measurements. Left: top view; middle: 3D view; right: average position error.

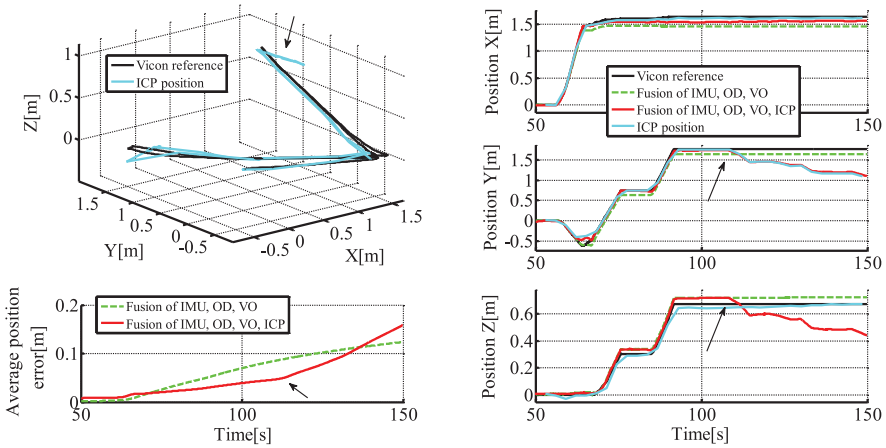


Figure 20. Trajectory estimates in the case of a moving obstacle in a reduced field of view. Solid black line: ground truth; solid red line: state estimate with all four modalities; dashed green line: state estimate excluding ICP measurements; cyan line: position estimated by ICP alone; black arrow: start of moving obstacle. Top left: 3D view; bottom left: average error as a function of time; right: x, y, and z coordinates as a function of time.

5.4.4. Moving Obstacle and Limited Laser Range

Unlike the cameras, laser range sensors are not sensitive to illumination conditions. On the other hand, they have a limited sensor range that can induce a lack of points in large environments. Close-range obstacles might then be the dominant cluster of points, and hence the ICP registration might converge to a wrong local minimum, following the motion of the obstacles.

To test this situation, we artificially limited the range of the laser range sensor to 2 m. This is similar to heavy smoke or dust scenarios that can arise in USAR conditions. This prevents the laser from observing the walls and the ceiling, which are usually the strongest cues for correct point cloud registration indoors.

Additionally, we used a large board to simulate a moving obstacle of significant size. This caused the ICP to drift, following the motion of the board.

Figure 20 shows the result of the filter compared to the ground truth. We can see that when the large obstacle starts to move, the estimate of the ICP drifts with it. As a

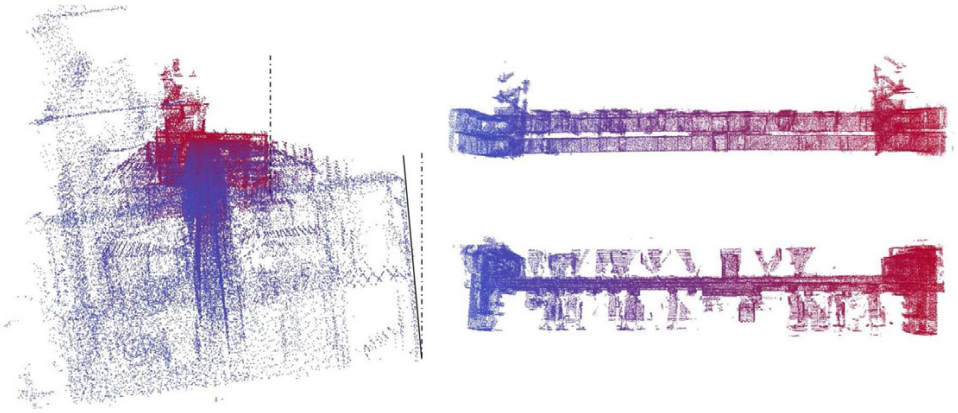


Figure 21. Deformed point cloud map created by ICP. The points are colored alongside the corridor from red (initial position) to blue. Left: front view; top right: side view; bottom right: top view.

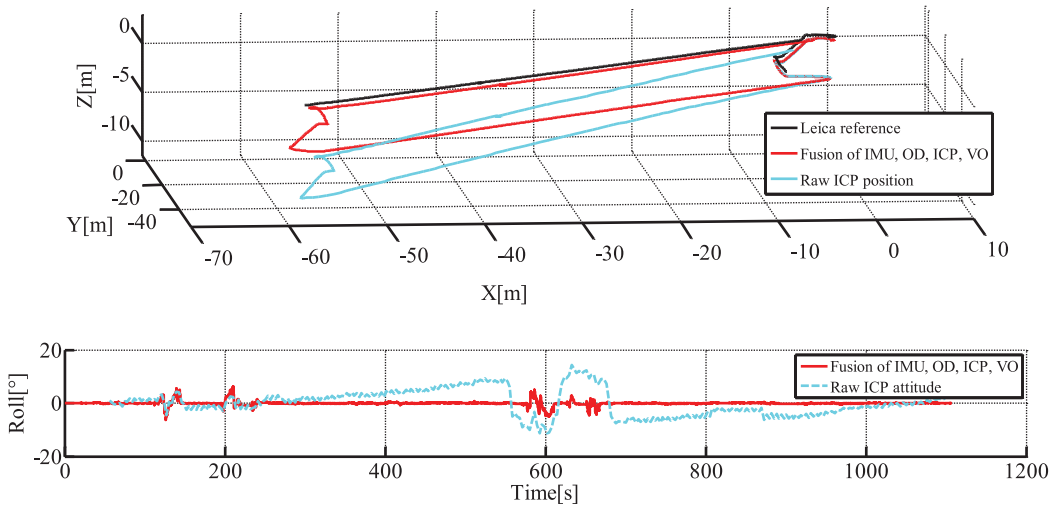


Figure 22. Trajectory estimates in the case of map deformation. Solid black line: ground truth; solid red line: state estimate with all four modalities; cyan line: position estimated by ICP alone. Top: side view; bottom: roll angle along the trajectory.

consequence, the whole filter drifts as well. This is analogous to the slippage situation, in which the ICP modality compensates for the combined estimate of the other three modalities. Using the omnicamera information not only as a visual compass but also as a complete visual odometry modality would probably allow us to differentiate between those two situations.

5.4.5. Map Deformation

As explained above, the ICP map is not globally optimized. This means that the map might have some large-scale deformations due to the accumulation of small errors. We were able to observe this particularly in a long corridor that we used to assess the impact of map deformation on the state estimate.

Figure 21 shows an instance of the deformed map. We drove along two superposed corridors over two floors. We can see that both ends of the corridor are not aligned: the ground plane of the blue end has a roll angle of several degrees compared to the red end. We used the theodolite system to acquire ground truth on the upper floor.

Figure 22 shows the impact of map deformation on the state estimate. The top graph shows that even if the ICP estimate is erroneous, the full filter maintains a correct, drift-free estimate. The bottom graph compares the estimate

of the roll angle between ICP only and the fusion. It clearly shows the drift in roll of the ICP estimate and the lack of impact it has on the fusion. The difference with previous failure case lies in the kind of drift. The drift of the roll angle can be compensated for by the IMU, especially the accelerometer. On the other hand, the drift in position of previous failure cases is not observable by the other modalities.

6. CONCLUSION

We designed and evaluated a multimodal data fusion system for state estimation of a mobile skid-steer robot intended for urban search and rescue missions. USAR missions often involve indoor and outdoor environments with challenging conditions such as slippage, moving obstacles, bad or changing light conditions, etc. To cope with such environments, our robot is equipped with both proprioceptive (IMU, tracks odometry) and exteroceptive (laser rangefinder, omnidirectional camera) sensors. We designed such a data fusion scheme in order to adequately include measurements from all four of these modalities with an order-of-magnitude difference in update frequency from 90 Hz to $\frac{1}{3}$ Hz.

We tested our algorithm on approximately 4.4 km of field tests (over more than 9 h of data) both indoors and outdoors. To ensure precise quantitative analysis, we recorded ground truth using either a Vicon motion capture system (indoors) or a Leica theodolite tracker (outdoors). In so doing, we proved that our scheme is a significant improvement upon standard approaches. Combining all four modalities—IMU, tracks odometry, visual odometry, and ICP-based localization—we achieved precision in the total distance driven of 1.2% error in the indoor environment and 1.4% error in the outdoor environment. Moreover, we characterized the reliability of our data fusion scheme against sensor failures. We designed failure case scenarios according to potential failures of each sensory modality that are likely to occur during real USAR missions. In the course of this testing, we evaluated robustness with respect to heavy slippage (odometry failure case), reduction of field of view of the omnicamera (visual odometry failure case), and reduction of the laser rangefinder together with large moving obstacles spoiling the created metric map (ICP-based localization failure case).

While our filter demonstrates good accuracy during our field tests and is robust against some of the failures expected in USAR, there is still room for improvement, namely the need for an automatic failure detection and resolution. Exploring different methods of detecting anomalous measurements and rejecting them in order to improve the overall performance is one of the ways, but it is currently left for future work. Furthermore, developing a visual odometry solution capable of also providing estimates of scaled translation is another topic for the future.

It is not surprising that combining more modalities yields greater precision. However, we were able to show that if such a rich multimodal system is well-designed, it will perform reasonably well even in cases in which other systems exploiting fewer modalities fail completely. We describe how to design such a system using the commonly used EKF. In this way, we contribute by proposing and comparing three different approaches to treat the ICP measurements, out of which the *trajectory approach* proved to perform best.

To contribute to the robotics community, we release our datasets used in this paper, including the ground truth measurements from the Vicon and Leica systems.

ACKNOWLEDGMENTS

The research presented here was supported by the European Union FP7 Programme under the NIFTi project (No. 247870; http://www.nifti.eu) and the TRADR project (No. 609763; http://www.tradr-project.eu). François Pomerleau was supported by a fellowship from the Fonds québécois de recherche sur la nature et les technologies (FQRNT). Vladimir Kubelka was supported by the Czech Science Foundation (Project Registration No. 14-13876S). We would like to thank the anonymous reviewers for their constructive suggestions that greatly improved the manuscript.

REFERENCES

Almeida, J., & Santos, V. M. (2013). Real time egomotion of a nonholonomic vehicle using lidar measurements. Journal of Field Robotics, 30(1), 129–141.

Anousaki, G., & Kyriakopoulos, K. J. (2004). A dead-reckoning scheme for skid-steered vehicles in outdoor environments. In Proceedings of the IEEE International Conference on Robotics and Automation (pp. 580–585).

Bachrach, A., Prentice, S., He, R., & Roy, N. (2011). RANGE—Robust autonomous navigation in GPS-denied environments. Journal of Field Robotics, 28(5), 644–666.

Barfoot, T., Stenning, B., Furgale, P., & McManus, C. (2012). Exploiting reusable paths in mobile robotics: Benefits and challenges for long-term autonomy. In Ninth Conference on Computer and Robot Vision (pp. 388–395).

Besl, P., & McKay, H. (1992). A method for registration of 3-D shapes. IEEE Transactions on Pattern Analysis and Machine Intelligence, 14(2), 239–256.

Breckenridge, W.G. (1999). Quaternions—Proposed standard conventions. Technical report, JPL.

Brodsky, T., Fermueller, C., & Aloimonos, Y. (1998). Directions of motion fields are hardly ever ambiguous. International Journal of Computer Vision, 26(1), 5–24.

Chen, Y., & Medioni, G. (1991). Object modeling by registration of multiple range images. In Proceedings of the IEEE International Conference on Robotics and Automation (pp. 2724–2729).

Chetverikov, D., Svirko, D., Stepanov, D., & Krsek, P. (2002). The trimmed iterative closest point algorithm. In Proceedings

of the 16th International Conference on Pattern Recognition (pp. 545–548).

Chiu, H.-P., Williams, S., Dellaert, F., Samarasekera, S., & Kumar, R. (2013). Robust vision-aided navigation using sliding-window factor graphs. In IEEE International Conference on Robotics and Automation (pp. 46–53).

Chowdhary, G., Johnson, E. N., Magree, D., Wu, A., & Shein, A. (2013). GPS-denied indoor and outdoor monocular vision aided navigation and control of unmanned aircraft. Journal of Field Robotics, 30(3), 415–438.

Civera, J., Grasa, O. G., Davison, A. J., & Montiel, J. M. M. (2010). 1-Point RANSAC for extended Kalman filtering: Application to real-time structure from motion and visual odometry. Journal of Field Robotics, 27(5), 609–631.

Dissanayake, G., Sukkarieh, S., Nebot, E., & Durrant-Whyte, H. (2001). The aiding of a low-cost strapdown inertial measurement unit using vehicle model constraints for land vehicle applications. IEEE Transactions on Robotics and Automation, 17(5), 731–747.

Ellekilde, L.-P., Huang, S., Miro, J. V., & Dissanayake, G. (2007). Dense 3D map construction for indoor search and rescue. Journal of Field Robotics, 24(1–2), 71–89.

Endo, D., Okada, Y., Nagatani, K., & Yoshida, K. (2007). Path following control for tracked vehicles based on slip-compensating odometry. In Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (pp. 2871–2876).

Fraundorfer, F., & Scaramuzza, D. (2012). Visual odometry: Part II: Matching, robustness, optimization, and applications. IEEE Robotics Automation Magazine, 19(2), 78–90.

Galben, G. (2011). New three-dimensional velocity motion model and composite odometry–inertial motion model for local autonomous navigation. IEEE Transactions on Vehicular Technology, 60(3), 771–781.

Jesus, F., & Ventura, R. (2012). Combining monocular and stereo vision in 6d-slam for the localization of a tracked wheel robot. In IEEE International Symposium on Safety, Security, and Rescue Robotics (pp. 1–6).

Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. Journal of Basic Engineering, 82(1), 34–45.

Kelly, J., Sibley, G., Barfoot, T., & Newman, P. (2012). Taking the long view: A report on two recent workshops on long-term autonomy. IEEE Robotics & Automation Magazine, 19(1), 109–111.

Kohlbrecher, S., Stryk, O. V., Meyer, J., & Klingauf, U. (2011). A flexible and scalable SLAM system with full 3d motion estimation. In IEEE International Symposium on Safety, Security, and Rescue Robotics (pp. 155–160).

Konolige, K., Agrawal, M., & Sola, J. (2011). Large-scale visual odometry for rough terrain. In Kaneko, M., and Nakamura, Y. (eds.), Robotics research, Vol. 66 of Springer Tracts in Advanced Robotics (pp. 201–212). Springer.

Kruijff, G. J. M., Janicek, M., Keshavdas, S., Larochelle, B., Zender, H., Smets, N. J. J. M., Mioch, T., Neerincx, M. A., van Diggelen, J., Colas, F., Liu, M., Pomerleau, F., Siegwart, R., Hlavac, V., Svoboda, T., Petricek, T., Reinstein, M., Zimmerman, K., Pirri, F., Gianni, M., Papadakis, P., Sinha, A., Balmer, P., Tomatis, N., Worst, R., Linder, T., Surmann, H., Tretyakov, V., Surmann, H., Corrao, S., Pratzler-Wanczura, S., & Sulk, M. (2012). Experience in system design for human-robot teaming in urban search and rescue. In Field and Service Robotics (pp. 1–14). Matsushima, Japan.

Kubelka, V., & Reinstein, M. (2012). Complementary filtering approach to orientation estimation using inertial sensors only. In IEEE International Conference on Robotics and Automation (pp. 599–605).

Kummerle, R., Grisetti, G., Strasdat, H., Konolige, K., & Burgard, W. (2011). g2o: A general framework for graph optimization. In IEEE International Conference on Robotics and Automation (pp. 3607–3613).

Lamon, P., & Siegwart, R. (2004). Inertial and 3D-odometry fusion in rough terrain—Towards real 3D navigation. In Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (pp. 1716–1721).

Li, H., & Hartley, R. (2006). Five-point motion estimation made easy. In 18th International Conference on Pattern Recognition (Vol. 1, pp. 630–633).

Ma, J., Susca, S., Bajracharya, M., Matthies, L., Malchano, M., & Wooden, D. (2012). Robust multi-sensor, day/night 6-dof pose estimation for a dynamic legged vehicle in gps-denied environments. In IEEE International Conference on Robotics and Automation (pp. 619–626).

McElhoe, B. A. (1966). An assessment of the navigation and course corrections for a manned flyby of Mars or Venus. IEEE Transactions on Aerospace and Electronic Systems, 2(4), 613–623.

Morales, Y., Carballo, A., Takeuchi, E., Aburadani, A., & Tsubouchi, T. (2009). Autonomous robot navigation in outdoor cluttered pedestrian walkways. Journal of Field Robotics, 26(8), 609–635.

Nagatani, K., Okada, Y., Tokunaga, N., Kiribayashi, S., Yoshida, K., Ohno, K., Takeuchi, E., Tadokoro, S., Akiyama, H., Noda, I., Yoshida, T., & Koyanagi, E. (2011). Multirobot exploration for search and rescue missions: A report on map building in robocuprescue 2009. Journal of Field Robotics, 28(3), 373–387.

Nemra, A., & Aouf, N. (2010). Robust INS/GPS sensor fusion for UAV localization using SDRE nonlinear filtering. IEEE Sensors Journal, 10(4), 789–798.

Nuchter, A., Lingemann, K., Hertzberg, J., & Surmann, H. (2007). 6D SLAM—3D mapping outdoor environments. Journal of Field Robotics, 24(8-9), 699–722.

Oskiper, T., Chiu, H.-P., Zhu, Z., Samarasekera, S., & Kumar, R. (2010). Multi-modal sensor fusion algorithm for ubiquitous infrastructure-free localization in vision-impaired environments. In 2010 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) (pp. 1513–1519).

Pomerleau, F., Colas, F., Siegwart, R., & Magnenat, S. (2013). Comparing ICP variants on real-world data sets. Autonomous Robots, 34(3), 133–148.

Reinstein, M., & Hoffmann, M. (2013). Dead reckoning in a dynamic quadruped robot based on multimodal

proprioceptive sensory information. IEEE Transactions on Robotics, 29(2), 563–571.

Reinstein, M., Kubelka, V., & Zimmermann, K. (2013). Terrain adaptive odometry for mobile skid-steer robots. In Proceedings of the IEEE International Robotics and Automation (ICRA) Conference (pp. 4706–4711).

Rodriguez, F. S. A., Fremont, V., & Bonnifait, P. (2009). An experiment of a 3D real-time robust visual odometry for intelligent vehicles. In Proceedings of the 12th International IEEE Conference on Intelligent Transportation Systems (pp. 1–6).

Rublee, E., Rabaud, V., Konolige, K., & Bradski, G. (2011). ORB: An efficient alternative to SIFT or SURF. In IEEE International Conference on Computer Vision (pp. 2564–2571).

Sakai, A., Tamura, Y., & Kuroda, Y. (2009). An efficient solution to 6DOF localization using unscented Kalman Filter for planetary rovers. In Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (pp. 4154–4159).

Savage, P. G. (1998). Strapdown inertial navigation integration algorithm design part 2: Velocity and position algorithms. Journal of Guidance, Control, and Dynamics, 21(2), 208–221.

Scaramuzza, D., & Fraundorfer, F. (2011). Visual odometry (tutorial). IEEE Robotics and Automation Magazine, 18(4), 80–92.

Shen, J., Tick, D., & Gans, N. (2011). Localization through fusion of discrete and continuous epipolar geometry with wheel and IMU odometry. In Proceedings of the American Control Conference (ACC) (pp. 1292–1298).

Smith, G. L., Schmidt, S. F., & McGee, L. A. (1962). Optimal filtering and linear prediction applied to a midcourse navigation system for the circumlunar mission. Technical report, U.S. Government Printing Office.

Sukumar, S. R., Bozdogan, H., Page, D. L., Koschan, A. F., & Abidi, M. A. (2007). Sensor selection using information complexity for multi-sensor mobile robot localization. In IEEE International Conference on Robotics and Automation (pp. 4158–4163).

Suzuki, T., Kitamura, M., Amano, Y., & Hashizume, T. (2010). 6-DOF localization for a mobile robot using outdoor 3D voxel maps. In Proceedings of the IEEE/RSJ International Intelligent Robots and Systems (IROS) Conference (pp. 5737–5743).

Svoboda, T., Pajdla, T., & Hlaváč, V. (1998). Motion estimation using central panoramic cameras. In IEEE International Conference on Intelligent Vehicles (pp. 335–340).

Tardif, J., Pavlidis, Y., & Daniilidis, K. (2008). Monocular visual odometry in urban environments using an omnidirectional camera. In IEEE/RSJ International Conference on Intelligent Robots and Systems (pp. 2531–2538).

Titterton, D. H., & Weston, J. L. (1997). Strapdown inertial navigation technology. Lavenham, UK: The Lavenham Press, Ltd.

Trawny, N., & Roumeliotis, S. I. (2005). Indirect Kalman filter for 3D attitude estimation—A tutorial for quaternion algebra. Technical report, University of Minnesota.

Van Loan, C. F. (1978). Computing integrals involving the matrix exponential. IEEE Transactions on Automatic Control, 23(3), 395–404.

Weiss, S. M. (2012). Vision based navigation for micro helicopters. Dissertation, ETH Zurich.

Yi, J., Zhang, J., Song, D., & Jayasuriya, S. (2007). IMU-based localization and slip estimation for skid-steered mobile robots. In Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (pp. 2845–2850).

Yoshida, T., Irie, K., Koyanagi, E., & Tomono, M. (2010). A sensor platform for outdoor navigation using gyro-assisted odometry and roundly-swinging 3D laser scanner. In Proceedings of the IEEE/RSJ International Intelligent Robots and Systems (IROS) Conference (pp. 1414–1420).

5. Improving multimodal data fusion for mobile robots by trajectory smoothing

This section presents the third publication EKF Smoothing [A.4] of this compiled Ph.D. thesis. It is a 9-page-long paper published in Robotics and Autonomous Systems journal. It extends findings of the previous publication Data Fusion [A.2] and compares them to an alternative state-of-the-art approach.

After **introducing** the reader to the problem of EKF-based localization data fusion system from the previous publication, contributions of this paper are stated. It is comparison of the *trajectory* approach with another possible technique which is applying a Kalman filter smoother. The other contribution is improving measurement model for visual odometry, which allows incorporating velocity information beside angular velocity information.

The **related work** section gives references to publications relevant to each sensory modality and to mobile robot localization in general. Kalman filter smoothers are another topic, we focus on their utilization in context of localization. Alternative techniques are cited as well, inverse filters and graph optimization methods.

Section **Smoother for multimodal data fusion** introduces our robotic platform and data fusion system briefly and focuses on the proposed *trajectory* approach to incorporating low-sampling-rate measurements, which we later compare to the Kalman filter smoother. The improved visual odometry measurement model is proposed as well as the way we apply the Rauch-Tung-Striebel smoother to the Extended Kalman filter.

The last part is the **Experimental evaluation**, which runs the localization system with smoother and with the *trajectory* approach and compares resulting accuracy. Experimental dataset overlaps with the one from Data Fusion [A.2], but includes additional experiments which include deformable terrain or running over swinging ramps. Observed results are discussed and summed up in the **Conclusions** section.

Robotics and Autonomous Systems 84 (2016) 88–96

Contents lists available at ScienceDirect

Robotics and Autonomous Systems

journal homepage: www.elsevier.com/locate/robot

Improving multimodal data fusion for mobile robots by trajectory smoothing

Vladimír Kubelka [a,b,*], Michal Reinstein [a], Tomáš Svoboda [a,b]

[a] Department of Cybernetics, Faculty of Electrical Engineering, Czech Technical University in Prague, Czech Republic
[b] Czech Institute of Informatics, Robotics, and Cybernetics, Czech Technical University in Prague, Czech Republic

HIGHLIGHTS

- The problem of fusing sensor modalities with significantly different sampling rates in a mobile robot localization system is addressed.
- A heuristic approach to include a low-rate position increment modality is proposed.
- The proposed approach is grounded with respect to a standard Rauch–Tung–Striebel smoother for the Kalman filter.
- Performance of the proposed approach is experimentally evaluated and selected fail-cases are discussed.

ARTICLE INFO

Article history:
Received 13 July 2015
Received in revised form
19 May 2016
Accepted 25 July 2016
Available online 2 August 2016

Keywords:
Field robots
Sensor fusion
Search and rescue robots

ABSTRACT

Localization of mobile robots is still an important topic, especially in case of dynamically changing, complex environments such as in Urban Search & Rescue (USAR). In this paper we aim for improving the reliability and precision of localization of our multimodal data fusion algorithm. Multimodal data fusion requires resolving several issues such as significantly different sampling frequencies of the individual modalities. We compare our proposed solution with the well-proven and popular Rauch–Tung–Striebel smoother for the Extended Kalman filter. Furthermore, we improve the precision of our data fusion by incorporating scale estimation for the visual modality.

1. Introduction

For successful deployment of mobile robots to complex dynamically changing environments, such as those typical for Urban Search & Rescue (USAR), reliable localization is crucial. In modern mobile robots, a popular solution lies in the combination of proprioceptive sensors, usually in form of an integrated Inertial Navigation System (INS), that captures the body dynamics at high rate, and an external source of aiding, using either vision [1] or range measurements [2]. Since most of the solutions are based on the well-proven Extended Kalman filter (EKF) [1,2], the state estimation architecture we designed for our platform (see Fig. 1) is based on the error state EKF framework as well.[1]

As we have shown in [3] (results of this work are summarized in Sections 3.1 and 3.2), performing data fusion of various modalities – such as in our case the inertial data, track odometry, visual odometry, and laser-based mapping – provides satisfactory results even when exposed to harsh environmental conditions, which can cause some of the modalities to fail. There is a number of well known problems connected with each named modality. First, the track odometry is strongly susceptible to high slippage, especially in skid-steer robots such as ours [4]. Second, it is the drift of the inertial sensors caused primarily by integrating the sensor noise, misalignment and instrumental errors. Third, the sensitivity to illumination and lack of scene texture influence the visual odometry performance [5]. And fourth, the laser-based mapping is sensitive to dynamic changes and to the overall geometric structure of the environment [6,7]. We addressed all these issues in [3] and introduced a failure-case methodology for evaluation of our multimodal data fusion. In this methodology we invoke challenging conditions that cause different modalities to fail on purpose and hence allow us to properly evaluate the robustness of localization.

However, our currently published results [3] raised a question that motivated us into a more in-depth research of the critical

* Corresponding author at: Czech Institute of Informatics, Robotics, and Cybernetics, Czech Technical University in Prague, Czech Republic.
E-mail addresses: kubelvla@fel.cvut.cz, kubelvla@fel.cvut.cz (V. Kubelka), reinstein.michal@fel.cvut.cz (M. Reinstein), svobodat@fel.cvut.cz (T. Svoboda).

[1] TRADR: Long-Term Human–Robot Teaming for Robot-Assisted Disaster Response www.tradr-project.eu.

http://dx.doi.org/10.1016/j.robot.2016.07.006
0921-8890/

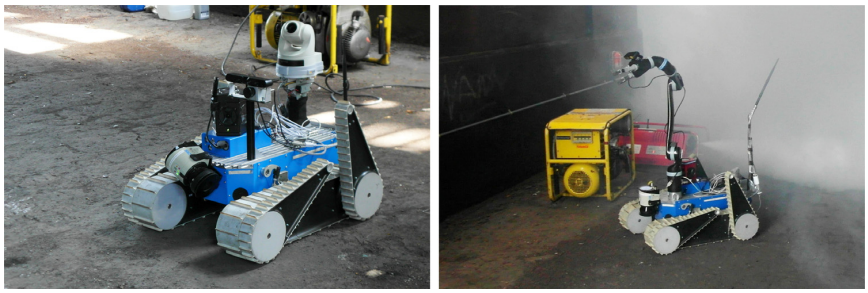


Fig. 1. TRADR[1] robotic platform at the USAR training site of the fire brigade of Dortmund, Germany. Robot sensor suite is highly configurable, includes Point Grey Ladybug3 omni-directional camera and SICK LMS-151 laser range-finder, can be extended further by various cameras (left) or a robotic arm (right).

issue of significantly different sampling frequencies of individual modalities. This is in general relevant to all multimodal data fusion algorithms, regardless a platform type. In this paper we therefore present our most recent results and compare them to the popular and commonly used standard smoother for Kalman filters. We have chosen the Rauch–Tung–Striebel smoother (RTS) [8] for the EKF as a representative of the Kalman smoothers family. By its definition, RTS best fits our data fusion scenario since it allows to recompute past position estimates based on information introduced by low-rate position increment measurements. We therefore exploited the RTS as benchmark for our multimodal data fusion architecture. For this purpose, we exploit our multimodal dataset[2] [3], which includes precise ground truth for both position and orientation obtained using a Vicon tracking system.

Our contribution is twofold. Due to experimental comparison and analysis, we were able to ground our novel approach to fusing multiple modalities at significantly different sampling rates with respect to the RTS smoother for EKF (described in Section 3.4). We hence offer our solution as an alternative to this popular RTS smoother, whether intended for robotics application or multimodal data fusion in general. Secondarily, with respect to our previous results, we improved the multimodal data fusion by incorporating velocity information from the visual odometry and resolved the scale problem for processing panoramic images (Section 3.3).

The paper is structured as follows: Section 2 introduces the related work, Section 3 sums up our previous work, describes our new proposed solutions and presents them in the context of smoothers for the EKF. Section 4 summarizes the experimental evaluation and Section 5 concludes the implications of our work.

2. Related work

Regarding the multi-modal data fusion, we built on our previous results described in [3], especially the design of the EKF error models [9–11]—even though the later work concerned a legged robot.

If long-term reliability and good accuracy are required, dead-reckoning solutions – such as those based on IMU and odometry – need other exteroceptive aiding modalities. In [12] it is shown that an IMU based dead reckoning system can be realized and successfully combined with the visual odometry to produce a reliable navigation system. We include visual odometry measurements into the EKF fusion scheme as well, yet directly in a form of angular and translational velocities computed by a more general implementation of visual odometry [5] designed for an omni-directional camera (note that in [12], the problem of tracking visual features is simplified by using a marker for planar homography).

Besides the visual odometry, another typical sensor for aiding is the laser range-finder. The laser range-finders are usually used for estimating vehicle motion by matching consecutive laser scans and thus creating a metric map of the environment [6,7]. Examples of successful deployment can be found for indoor – without IMU but combined with vision [13] – as well as for outdoor—relying only on the IMU [2]. The most popular approach of scan matching is based on the Iterative Closest Point (ICP) algorithm, which was first proposed by [14,15]. Later, [16] proposed a 6D Simultaneous Localization and Mapping (SLAM) system relying primarily on the ICP. Work of [17] proposed a localization system combining a 2D laser SLAM with a 3D IMU/odometry-based navigation subsystem. Contrary to the later publications realized in the context of SLAM, we only consider the output of the ICP algorithm[3] as a local pose measurements—similarly as with the visual odometry, we treat the laser localization module as a *velocity sensor*.

Solutions exploiting the EKF for fusing the dead reckoning with exteroceptive sensors are very popular [1,2,19–21]; our fusion scheme is based on the EKF as well. Still, a number of problems arise in multimodal data fusion. The problem of utilizing several sensors for localization, which may provide contradictory measurements, is discussed in [22]. The authors use Bayes filters to estimate sensor measurement uncertainty and hence evaluate the sensor validity. We separately addressed this problem in [10], where we utilized machine-learning techniques to detect anomalous measurements.

Since we aim for grounding our approach with respect to the smoothers for Kalman filters (in order to smooth the trajectory estimates), we have chosen the well established RTS [8] for the EKF as our base reference for benchmarking. Smoothers like RTS are well proven in the context of localization. In [23] a network of time-of-flight *Cricket* sensors provide measurements with a slight delay; the authors utilize an interacting multiple model fixed lag smoother to incorporate these delayed measurements. In [24] indoor localization problem is used to demonstrate properties of a smoother for the Unscented Kalman Filter. And finally in [25], the RTS smoother is actually utilized for the SLAM problem. Smoothing in Kalman filtering can be applied to wide range of problems, e.g. work of [26] applies the RTS to improve state estimation of a dynamic power system.

Several modifications of the RTS smoother have been proposed [27–29] since its original publication [8]. They mainly aim on better numerical stability and performance of the filter when deployed on computers with limited precision of number representation. We compare our algorithm to the RTS smoother using

[2] This dataset has been already released to the robotics community at https://sites.google.com/site/kubelvla/public-datasets/nifti-zurich-2013.

[3] We use the *libpointmatcher* implementation of the ICP algorithm [18].

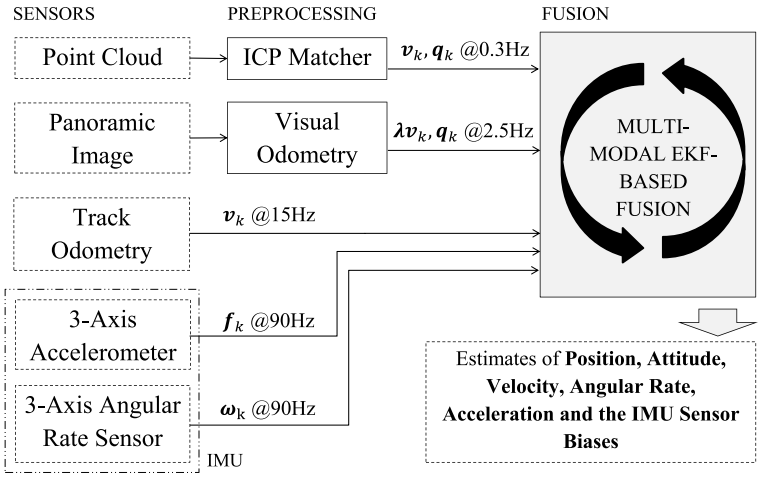


Fig. 2. Schematics of the multi-modal fusion algorithm.

64-bit double precision number representation and thus we take the liberty to use the original RTS formulation from [8].

Apart from the Kalman filtering frameworks, factor-graph-based approaches have recently gained popularity. They elegantly describe estimation problem by a bipartite graph composed of variable and factor nodes representing system states and measurements (application for localization is presented e.g. in [30,31]). Since batch optimization of the whole factor graph (which is smoothing over all past estimates) can be costly, work of [32] presents the *iSAM2* algorithm that allows incremental optimization of the factor graph. This approach is especially beneficial in the case of SLAM. Estimating the position of observed landmarks together with the position of camera or other sensor makes the problem highly non-linear. However, in the case of fusing velocity or angular rates for on-line state estimation without performing SLAM, the EKF framework provides comparable performance to the factor graph approach [33].

Another approach that implicitly involves smoothing are inverse filters for visual–inertial systems [34,35]. These filters offer straightforward mechanism for re-linearizing measurements and re-processing visual observations that are within a defined optimization window. Work of [36] further provides a smoother that combines benefits of numerical stability of the EKF and re-linearization capability of inverse filters. Nevertheless, complete redesign to a tightly coupled systems is out of scope of our work.

3. Smoother for multimodal data fusion

3.1. Platform and sensors

For the estimation of position and orientation, our localization algorithm fuses four sensor modalities: the inertial data from an IMU and track odometry, laser range-finder and visual odometries. As Fig. 2 demonstrates, each of these modalities generates measurements with different sampling rates. Proprioceptive sensors (*Xsens MTi-G* IMU and track velocity encoders) provide high-sapling-rate measurements, yet they are prone to drift caused by measurement noise integration and effects of track slippage. The *Xsens MTi-G* IMU unit offers internal EKF-based attitude estimation but we choose to utilize its raw measurements instead (since the fusion with other measurements allows better attitude estimation). The IMU is connected to the system via standard serial link (COM port), we neglect delay between the actual measurement and read-out from the serial port buffer; we time-stamp it by the current system time. The IMU sampling rate is approximately 90 Hz.

Visual and laser range-finder odometries rely on omni-directional images captured by the *Point Grey Ladybug3* camera and on range data acquired by the *SICK LMS-151* sensor respectively (the native functionality of the laser range finder is 2D scans; it is rotated by a servo drive to obtain 3D scans). The *Ladybug3* camera is connected via Firewire 400 port; which is saturated by high amount of the image data thus leading to reading delays. Nevertheless, the camera time-stamps the images by the true time of the capture. The laser scanner is connected to the system by ethernet link; the raw data are not time-stamped. We introduce a time correction for the delay between measuring and the data read-out (a small fraction of a second) that was determined empirically.

The 3D scan can be generated even as the robot moves. For this purpose, position is estimated from track odometry and IMU measurements—resulting 3D scans are accurate enough for the ICP algorithm to converge [18]. Since our target applications to USAR do not allow reasonable use of GPS or even magnetometer, we omit these modalities in our multimodal data fusion.

3.2. Multimodal data fusion framework

Output of the four modalities is preprocessed or directly enters the EKF-based[4] fusion algorithm that estimates – among others – the 3D position and orientation of the robot. Fig. 2 presents the outline of the fusion system where $\boldsymbol{v}_k$ and $\boldsymbol{\omega}_k$ stand for linear and angular velocities, $\boldsymbol{f}_k$ is specific force measured by the accelerometers. Attitude provided by the laser range-finder and visual odometries is represented by quaternion $\boldsymbol{q}_k$; note that visual odometry indicates velocity multiplied by an unknown scale λ, which has to be estimated by the EKF fusion algorithm. As explained in [3], in order to design a modular solution, we treat all the modalities as velocity sensors.

The exteroceptive sensors generate measurements that require preprocessing in order to obtain information about translation and rotation of the robot. The laser range-finder creates 3D scans

[4] We considered using the Unscented Kalman filter as well but according to our experience from work [9], the UKF does not improve the fusion performance since the model of the robot dynamics is only an approximation of the true state.

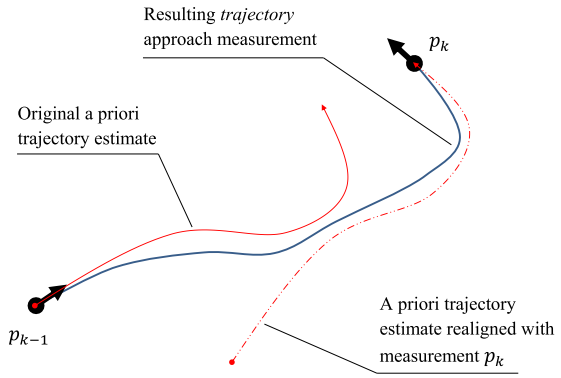


Fig. 3. The trajectory approach [3]. Trajectory estimated from all modalities except the low-sampling-rate sensor (solid red line) is combined with position information (p_k, p_{k-1}) from the low-sampling-rate sensor (laser odometry). The new trajectory (thick blue) is a weighted average of the original trajectory and its aligned duplicate (red dot-dashed line). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

of the environment. We use *libpointmatcher* implementation [18] of the ICP algorithm to determine a rigid transformation that would match two successive point clouds together. Resulting transformation reflects movement of the robot between the two 3D scans. According to our experience, this estimate is reliable in most environments the robot operates in. Its main limitation – the low sampling rate of one scan per 3 s – is given by the speed of rotation of the laser range-finder.

In general, there are 3 different approaches to incorporate these low-sampling-rate measurements into the fusion algorithm [3]. We consider two of them to be state-of-the-art common practice. The first expresses the measured translation as an increment in position [37], the second as a constant velocity observed between two consecutive measurements [38]. Both solutions did not provide satisfying results when applied to measurements with such a low sampling rate. Therefore, we proposed a heuristic approach that overcomes some of the problems cased by the sampling rate. We call it the *trajectory* approach since it utilizes the shape of the trajectory estimated by the remaining modalities to generate velocity measurements [3].

The basic idea behind the *trajectory* approach is that while the position and orientation estimates deteriorate with time and distance traveled, locally, the shape of the trajectory resembles the true one. We choose two time instances – two consecutive laser odometry measurements – and claim that if we align a priori trajectory estimate (the one originating from other modalities) with these measurements, we obtain locally accurate trajectories around the two measurements. See Fig. 3 that depicts the aligned a priori and the resulting trajectories. We obtain the resulting trajectory by weighted average of the original one and its realigned duplicate.

This heuristic approach allows us to generate velocity measurements that better fit into the EKF fusion scheme. Also, it is possible to further apply non-holonomic constraints on the velocity measurements, which help to reduce the drift in the vertical axis mainly caused by slight inaccuracies in the pitch angle indicated by the laser odometry.

3.3. Scale-dependent velocity information from visual odometry

Relative motion of the robot is also measured by the visual odometry (VO). It benefits from the omni-directional camera observing the whole scene (up to occlusions caused by the robot body) and thus it is not losing observed image features while turning. The omni-directional camera consists of several perspective cameras built in a way that purposively puts individual camera centers very close to each other and it is modeled as a central omni-directional camera. Therefore, scale of the scene cannot be computed directly and it is up to the fusion algorithm to estimate it. The estimated scale then affects the velocity indicated by the VO. Since the VO scale has not been the part of the original fusion algorithm we proposed in [3] (and thus only VO attitude corrections could be utilized), we augment it to the state vector as follows:

$$\boldsymbol{x} = \begin{bmatrix} \boldsymbol{p}_N & \boldsymbol{q}_N^R & \boldsymbol{v}_R & \boldsymbol{\omega}_R & \boldsymbol{f}_R & \boldsymbol{b}_{\omega,I} & \boldsymbol{b}_{f,I} & \lambda \end{bmatrix}^T \tag{1}$$

where $\boldsymbol{p}_N$ is position of the robot in the N-frame (Navigation/world frame), $\boldsymbol{q}_N^R$ is unit quaternion representing attitude, $\boldsymbol{v}_R$ is velocity expressed in the R-frame (Robot/body frame), $\boldsymbol{\omega}_R$ is angular rate, $\boldsymbol{f}_R$ is specific force [39], $\boldsymbol{b}_{\omega,I}$ and $\boldsymbol{b}_{f,I}$ are accelerometer and angular rate sensor IMU-specific biases expressed in the I-frame (Inertial frame). λ is the VO scale. Since we implemented the EKF in the error-state fashion (compare standard EKF equations with the error-state EKF in Fig. 4), the system state is accompanied by the corresponding error state $\Delta\boldsymbol{x}$:

$$\Delta\boldsymbol{x} = [\Delta\boldsymbol{p}_N \quad \delta\boldsymbol{\theta} \quad \Delta\boldsymbol{v}_R \quad \Delta\boldsymbol{\omega}_R \quad \Delta\boldsymbol{f}_R \quad \Delta\boldsymbol{b}_{\omega,I} \quad \Delta\boldsymbol{b}_{f,I} \quad \Delta\lambda]^T. \tag{2}$$

Measurement model for velocity indicated by VO is then

$$\boldsymbol{y}_{v,VO} = \lambda\boldsymbol{v}_R + \boldsymbol{m}_{v,VO} \tag{3}$$

where $\boldsymbol{m}_{v,VO}$ is measurement noise associated with this sensor modality. For the error-state EKF, we express associated measurement residual, i.e. how the error state contributes to the observed discrepancy between expected measurement and the actual one:

$$\begin{aligned} \boldsymbol{y}_{v,VO} - \hat{\boldsymbol{y}}_{v,VO} &= \Delta\boldsymbol{y}_{v,VO} \\ &= (\hat{\lambda} + \Delta\lambda)(\hat{\boldsymbol{v}}_R + \Delta\boldsymbol{v}_R) - \hat{\lambda}\hat{\boldsymbol{v}}_R + \boldsymbol{m}_{v,VO} \\ &\approx \hat{\boldsymbol{v}}_R\Delta\lambda + \hat{\lambda}\Delta\boldsymbol{v}_R + \boldsymbol{m}_{v,VO} \end{aligned} \tag{4}$$

where all the *hat* symbols stand for expected values obtained by the system state propagation in time. From the result of (4), appropriate rows of the measurement matrix H can be constructed; vector of all measurement residuals is then expressed as

$$\Delta\boldsymbol{y} = H\Delta\boldsymbol{x} + R \tag{5}$$

where R is the measurement noise matrix. Since the scale λ is not known, it is initialized to one and left to be estimated.

Before the value of scale of the VO converges, invalid corrections of the velocity v_R are propagated into the system state. Similarly, erroneous track odometry measurements (e.g. in case of slippage) will eventually propagate into the estimate of λ. Therefore, to be able to utilize the visual odometry in a reasonable manner, we propose a modification that separates the λ estimation phase from the VO aiding phase by omitting either the $\hat{\lambda}\Delta\boldsymbol{v}_R$ term or the $\hat{\boldsymbol{v}}_R\Delta\lambda$ term from (4). In the first case, only λ is estimated without any v_R corrections being introduced. Based on the current EKF filter setting, we empirically determined it takes approximately 1.5 m traveled by the robot for the λ to converge to a stable value (travel speed of the robot is in average 0.4 m/s). After this nominal distance, we switch the terms, so from that point, λ is constant and every measurement residual is propagated into v_R correction. Of course, this approach expects the visual odometry to maintain constant scale, which may not always be the case.[5]

[5] If the scale does not stay constant and anomalous VO measurements are detected by the standard χ^2 test or one of approaches proposed in [10], λ can be re-initialized again.

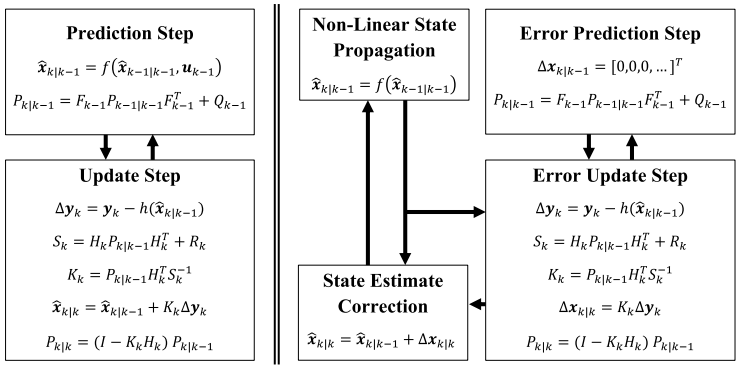


Fig. 4. Comparison of the standard (left) and error-state (right) Extended Kalman filter computational scheme. Note that the system matrix F_k differs between the two filter implementations, it describes either dynamics of the system state or of the error of the system state [3]. Both F_k and H_k (measurement matrix) are linear approximations of the general nonlinear functions describing the system. (Figure reprinted from [3].)

3.4. RTS smoother for the EKF

Since the main goal of this paper is to compare our *trajectory* approach with the combination of the standard state-of-the art *incremental position* approach with a smoother for KF, we introduce here the way we utilize the Rauch–Tung–Striebel smoother [8] with our error state EKF implementation.

By smoothing is understood estimating the state based on both past and future measurements. In general, smoothing can be used when a small delay in estimation can be tolerated or in case of post-processing. Standard RTS smoother is a two-pass filter which in the forward pass does not differ from the Kalman filter. The smoothing is then performed in the backward pass, given by the following equations. By subscript $[s]$, we denote the smoothed state; the last sample N is considered smoothed as it is:

$$\boldsymbol{x}_{[s],N} = \boldsymbol{x}_{N|N} \tag{6}$$

and the rest is smoothed using these recursive equations, where P is the state covariance matrix and subscripts $k+1|k$ and $k|k$ stand for a priori (at time $k+1$) and a posteriori (at time k) estimates of the state and its covariance:

$$\boldsymbol{x}_{[s],k} = \boldsymbol{x}_{k|k} + A_k \left(\boldsymbol{x}_{[s],k+1} - \boldsymbol{x}_{k+1|k}\right) \tag{7}$$

$$P_{[s],k} = P_{k|k} + A_k \left(P_{[s],k+1} - P_{k+1|k}\right) A_k^T \tag{8}$$

where the gain A_k is obtained from covariance matrices P and the system matrix F[6] as follows

$$A_k = P_{k|k} F_k^T P_{k+1|k}^{-1}. \tag{9}$$

In case of the error-state EKF, the procedure is very similar, only (7) has to be interpreted in terms of the error state. The term $\left(\boldsymbol{x}_{[s],k+1} - \boldsymbol{x}_{k+1|k}\right)$ has to follow the error state definition, which involves subtraction in all terms but the quaternion $\boldsymbol{q}$—rotation vector $\delta\boldsymbol{\theta}$ is used (2) instead of the equivalent quaternion expression

$$\delta\boldsymbol{q}_k = \boldsymbol{q}_{[s],k+1} \otimes \boldsymbol{q}_{k+1|k}^{-1} \tag{10}$$

where $\otimes$ stands for quaternion multiplication (see [40]). $\delta\boldsymbol{\theta}$ is then two times the vector part of $\delta\boldsymbol{q}$. Similarly, the following summation operation

$$\boldsymbol{x}_{[s],k} = \boldsymbol{x}_{k|k} + A_k \Delta\boldsymbol{x}_k \tag{11}$$

is not defined; the correction has to be applied following the error state EKF definition; i.e. summation in all terms but the quaternion, where the rotation vector $\delta\boldsymbol{\theta}$ is expressed as quaternion ($\delta\boldsymbol{q} = [0.5\delta\boldsymbol{\theta}^T\ 1]^T$) and multiplied by its counterpart in $\boldsymbol{x}_{k|k}$:

$$\boldsymbol{q}_{[s],k} = \boldsymbol{q}_{k|k} \otimes \delta\boldsymbol{q}_{k|k}. \tag{12}$$

Since the *trajectory* approach operates on states between two consecutive laser odometry measurements, we smooth the same states by the RTS in the case of the *incremental position* approach for our comparison.

4. Experimental evaluation

4.1. Dataset description

For the purpose of experimental evaluation, we used our dataset that we already released to the community.[7] It consists of experiments performed both outdoors and indoors, under reasonably challenging conditions—track slippages that occur naturally with the robot movement, traversing stairs, slopes, gravel etc. We also performed series of experiments, where we deliberately diminished sensors performance (covered part of the omni-directional camera, scene over-exposure, artificially limited laser range-finder range) or simulated rough conditions (slippery slopes, deformable and unstable surfaces, hitting obstacles; see Fig. 5). The complete dataset contains approximately 3.2 km experiments with ground truth. To be able to evaluate performance, indoor experiments (total length 1.6 km) were tracked by the Vicon system. With the ground truth, we can evaluate localization error in any given time yet such a metric is difficult to interpret, mainly because the error grows with distance traveled. To overcome this problem, we normalize the accumulated error by the distance and call it *average position error*:

$$e_{avg}(l) = \frac{\sum_{i=1}^{l} \|\boldsymbol{p}_i - \boldsymbol{p}_{ref,i}\|}{l} \tag{13}$$

where l is the position sample index. To improve legibility of this metric in plots, we express the e_{avg} as a function of time

$$e'_{avg}(t) = e_{avg}(l(t)) \tag{14}$$

where $l(t)$ simply maps time t to the corresponding sample l.

[6] For the full definition of the proposed state space model, the error model and its linearization see [3].

[7] [3,41], the dataset is recorded in a ROS *bagfile* format. Precise ground-truth position reference as captured by Vicon system is included.

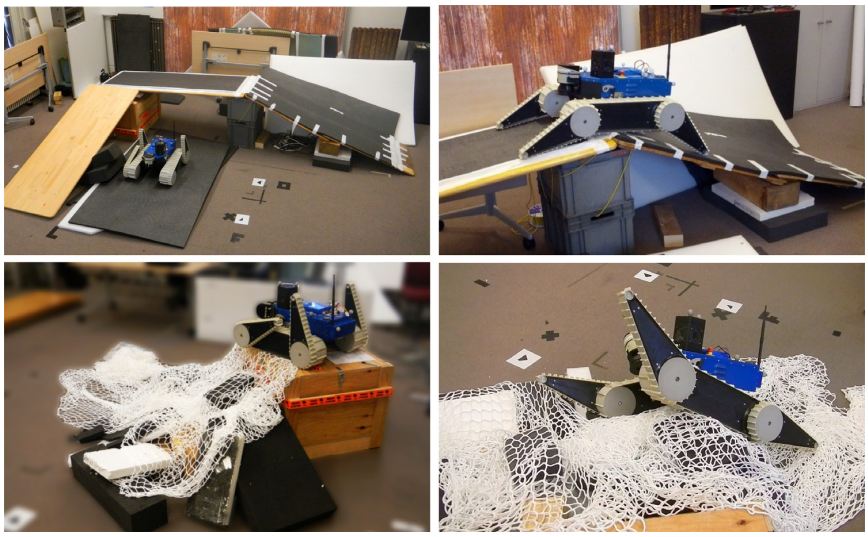


Fig. 5. Examples of obstacles traversed during indoor localization experiments (Vicon tracking system as a ground truth). Top images: Ramp with a slippery surface (lacquered wood) and a non-slippery one (soft rubber). Bottom images: Pile of soft plastic foam, polystyrene and a nylon net (the wooden box served as a side support for the pile).

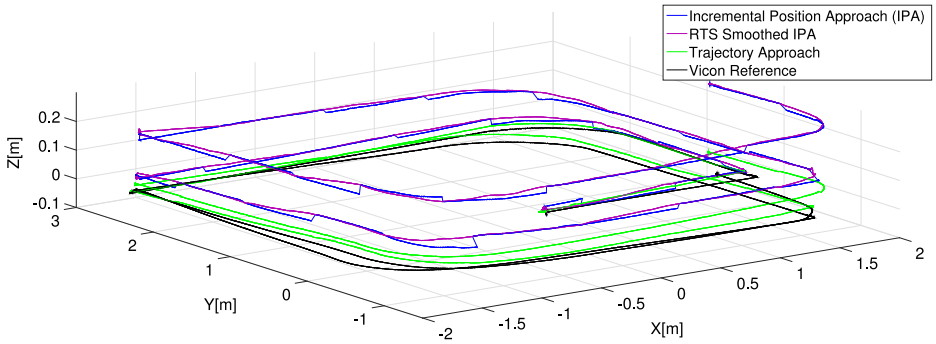


Fig. 6. Comparison of the discussed approaches in an experiment with nominal conditions—robot was navigated in a square-shaped pattern on a leveled floor.

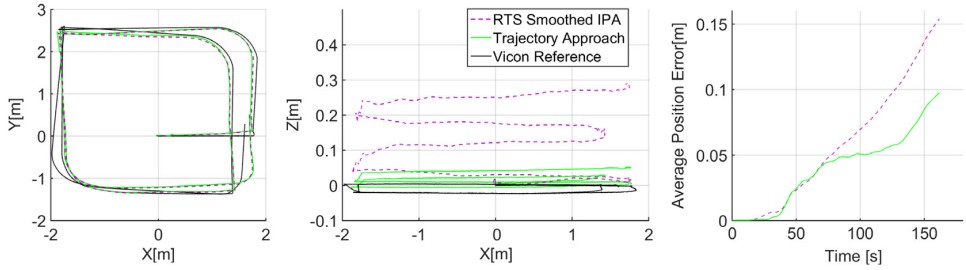


Fig. 7. Effect of drift in the Z-axis for the RTS smoothed incremental position approach (IPA) compared to the trajectory approach (same experiment as in Fig. 6).

4.2. Evaluation

We distinguish between experiments with nominal conditions and special fail-cases, where we deteriorated performance of selected sensor modalities or even caused their failure. Yet to properly test and compare the RTS smoothed *incremental position* approach with its *trajectory* approach counterpart (we are interested in the impact of these approaches to artifacts caused by erratic measurements), we include these fail-case experiments to the overall statistics.

In this subsection, we provide several examples of nominal and fail-case experiments (Figs. 6–10) that demonstrate behavior of both evaluated approaches. When navigating over flat surfaces (Fig. 6, 7 or 8) and inclined ramps without major slippage (Fig. 9) the trajectory approach allows to generate hi-rate velocity measurements from low-rate position increment measurements. These can be corrected by non-holonomic constraints that compensate effect of improper attitude estimation on the side of the low-rate odometry. Inaccurate attitude introduces errors when expressing measured velocity in the body coordinate frame. These

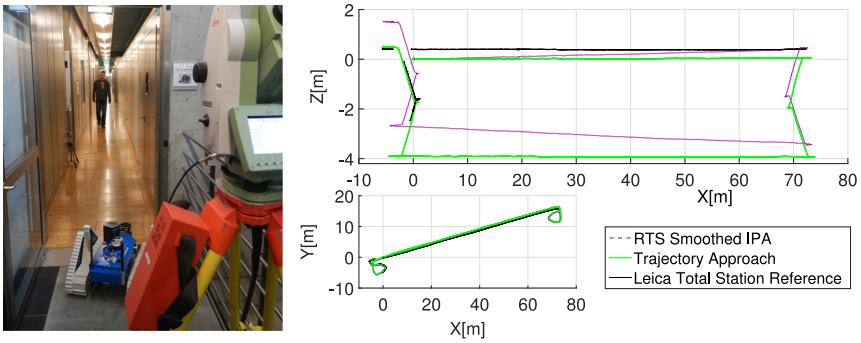


Fig. 8. Effect of the discussed approaches in a long narrow corridor experiment spanning two floors. Non-holonomic constraints in measured velocity (allowed by the trajectory approach) reduced the Z-axis drift. Note that the reference system was available only on the second floor and the offset between the reference and the trajectory was caused by tracking a prism on a 50 cm pole attached to the robot.

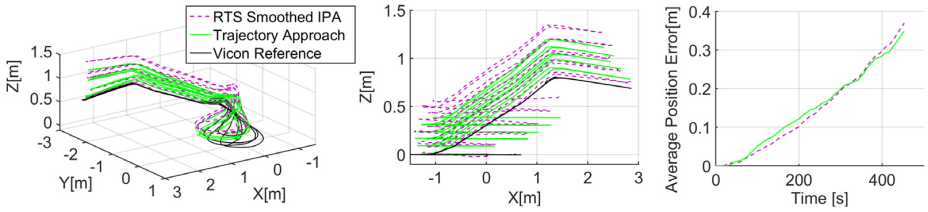


Fig. 9. A ramp-climbing fail-case experiment example. The robot repeatedly climbed an inclined ramp (see Fig. 5, top right). With each pass, error in the Z coordinate estimate increases because of inaccurate readings of the tracked odometry when rolling onto the inclined ramp and pivoting over its top.

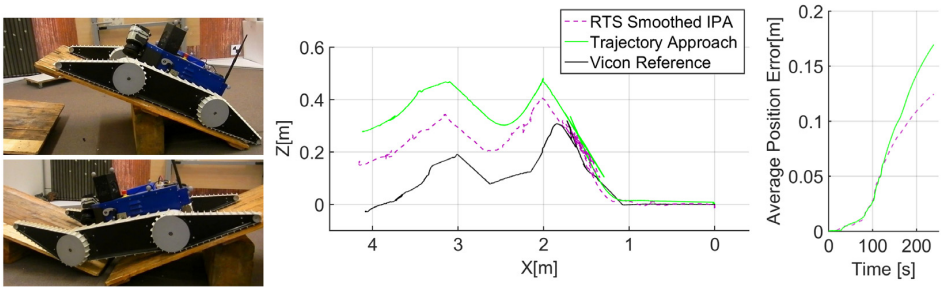


Fig. 10. An extreme case with slippage and swinging pallets. The pivoting motion of the pallet lying on the wooden block violates the non-holonomic constraints used in the trajectory approach. That results in worse position estimate than the smoothed incr. position approach.

errors – in our case – lead to drift in the Z-axis (vertical). The constraint we apply assures that the generated hi-rate velocity vector is parallel with the X–Y plane of the body coordinate frame of the robot. As the plots of results evaluated at nominal conditions demonstrate, our constraint of velocity measurement suppresses the drift in the vertical axis.

First example of experiments in Fig. 6, 7 compares the discussed approaches on a basic square trajectory. Both the incremental position approach and its RTS smoothed version show drift in the Z-axis, which is suppressed by the non-holonomic constraint in the case of the trajectory approach. The same effect of the trajectory approach is shown in a long corridor experiment in Fig. 8. The reference system was a Leica Total Station.[8]

As the motion of the robot becomes more complex, localization drift increases. This is shown in Fig. 9. In this experiment, robot repeatedly approached, climbed and descended an inclined ramp (see Fig. 5, top right). Each time the robot rolled onto the ramp and each time it pivoted over its top edge small localization error accumulated. In this case, the non-holonomic constraints diminished drift on flat parts of the trajectory but did not help when the robot passed over the edges and thus the localization accuracy of the two approaches was comparable.

An extreme fail-case of violating assumptions laid by the trajectory approach is demonstrated in the next example in Fig. 10. The robot traversed two seesaws made of wooden pallets and large wooden blocks serving as a pivot. At some point of the traversal, the pallet together with the robot pivoted over while performing rotational and translational motion. Both of motion components were sensed by the inertial measurement unit. Rotational part was estimated without any problems, however, the translation was based on data from visual odometry, IMU acceleration and track odometry which provided contradictory measurements. This resulted into incorrect localization for both approaches; non-holonomic constraints were violated.

The last presented fail-case related to the low-rate ICP odometry is depicted in Fig. 11. In this experiment, we intentionally

[8] Theodolite with a laser range finder which tracked a prism attached to the robot—the reference is thus available on the second floor of the two-floor trajectory. There is also cca. 0.5 m offset from the robot origin because of the way of attaching prism to the robot.

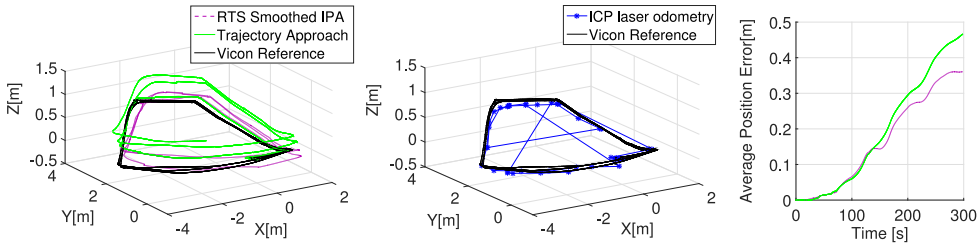


Fig. 11. An ICP laser odometry outage fail-case. The raw ICP measurements are shown in the middle plot (blue stars are the points of the ICP measurements); the nominal rate is 0.3 Hz however in this experiment, we switched laser for longer periods to simulate laser failures. The robot path also included the slippery ramp that caused invalid velocity measurements from track odometry. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 1
Comparison of the trajectory and smoothed incremental position approach in terms of average position error.

Approach	Normal conditions (777 m,7218 s)	+ Hard conditions (2174 m,15 931 s)
Incremental position	0.08|**0.12**|0.22	0.10|**0.18**|0.39
Incr. pos. with smoothing	0.08|**0.12**|0.21	0.10|**0.19**|0.38
Trajectory approach	0.11|**0.14**|0.24	0.12|**0.19**|0.36
	lower|**median**|higher	quartile (m)

paused laser range finder for up to 45 s simulating ICP odometry outages. During these outages, the robot passed over slippery and non-slippery ramp inducing track odometry errors discussed in previous experiment examples. It is necessary to comment on the middle plot in Fig. 11: Compared to our fused localization output, the bare ICP odometry yields minimal error. That is expected in a laboratory where it has perfect conditions for localization—six large planes of walls, ceiling and floor. This advantage does not hold in general, ICP odometry tends to perform poorly in uniform tube-shaped environments (e.g. long corridors). Nevertheless, comparison of the smoothed incremental position approach and the trajectory approach favors the first one. While slippage does not usually significantly deteriorate the trajectory approach performance, the long outages lead to poor trajectory reconstruction (Fig. 3), especially in the Z-axis leading to stronger drift.[9]

4.3. Comparative analysis

In Table 1, we evaluate performance of the approaches on the indoor set of experiments—the Vicon reference system provides higher sampling rate and thus allows us to evaluate overall accuracy as well as details of the estimated robot trajectory. The first column contains the set of experiments with nominal conditions; in the second column, the fail-case experiments are also included. The results show that the overall accuracy is comparable—the median values of average position error of each approach is well within lower and higher quartiles of the other approaches. We choose to express the results by lower and higher quartile and median since our metric value is not normally distributed. Including the fail-cases increases the median of the average position error, as expected.

4.4. Discussion

With respect to the presented results, obvious question arises: which approach should be chosen? On flat or reasonably inclined surfaces without rough transitions, trajectory approach corrects localization drift (predominantly in the Z-axis). It is the non-holonomic constraint on ICP odometry velocity measurements that mostly contributes to this correction. On rough and deformable terrain and when pivoting over obstacle edges, the non-holonomic constraint does not hold anymore and therefore, the standard incremental position approach performs similarly to our trajectory approach. In the Vicon-referenced part of our dataset, the dominant part is experiments where the robot crosses ramps and other obstacles. Therefore, the comparison of the approaches in Table 1 indicates similar performance. The choice of the approach therefore depends on the expected conditions of the robot deployment. If a significant part of trajectories driven by the robot contains large flat areas (structured environment in general), the localization can benefit from non-holonomic constraints implemented in the proposed trajectory approach.

Smoothing the position estimates (either by adding the RTS smoother into the incremental position approach or using the trajectory approach) can only improve estimates backwards into history. Closed-loop controllers will be negatively influenced by the delay, but there may be systems where the delay does not matter (typically registration of various sensor measurements with world coordinate frame).

5. Conclusions

We have improved our multimodal data fusion by incorporating velocity measurements obtained from monocular visual odometry, whose scale is in principle unknown. We have compared our *trajectory* approach with the combination of a state-of-the-art approach and the Rauch–Tung–Striebel smoother that was modified for the error-state EKF. We evaluated its performance on a set of experiments (over 2 km of distance traveled) designed to imitate USAR mission conditions—even those that cause some of the sensor modalities to fail.

The results have shown that both approaches are comparable in terms of average position error due to challenging composition of the dataset, yet the *trajectory* approach clearly outperforms the standard incremental position approach with RTS under nominal conditions. We further investigate conditions that are not nominal in [10], where we propose extension of the EKF fusion framework by an anomaly detection algorithm that allows us to cope with these conditions.

Acknowledgments

The research presented here was supported by the European Union FP7 Programme under the TRADR project (No. FP7-ICT-609763; http://www.tradr-project.eu), by the Czech Science Foundation (Project Registration No. GA14-13876S) and by the Grant

[9] In this case, however, the fusion system could be easily modified to detect ICP odometry outages and accommodate appropriately.

Agency of the CTU Prague under Project SGS15/081/OHK3/1T/13. We would like to thank F. Colas, F. Pomerleau and L. Oswald for their valuable comments and advices and for their assistance with creation of the experimental dataset.

References

[1] G. Chowdhary, E.N. Johnson, D. Magree, A. Wu, A. Shein, GPS-denied indoor and outdoor monocular vision aided navigation and control of unmanned aircraft, J. Field Robot. 30 (3) (2013) 415–438. URL http://dx.doi.org/10.1002/rob.21454.
[2] A. Bachrach, S. Prentice, R. He, N. Roy, RANGE—Robust autonomous navigation in GPS-denied environments, J. Field Robot. 28 (5) (2011) 644–666.
[3] V. Kubelka, L. Oswald, F. Pomerleau, F. Colas, T. Svoboda, M. Reinstein, Robust data fusion of multimodal sensory information for mobile robots, J. Field Robot. 32 (4) (2015) 447–473. URL http://dx.doi.org/10.1002/rob.21535.
[4] D. Endo, Y. Okada, K. Nagatani, K. Yoshida, Path following control for tracked vehicles based on slip-compensating odometry, in: Proc. IEEE/RSJ Int. Conf. Intelligent Robots and Systems IROS 2007, 2007, pp. 2871–2876. http://dx.doi.org/10.1109/IROS.2007.4399228.
[5] D. Scaramuzza, F. Fraundorfer, Visual odometry [tutorial], IEEE Robot. Autom. Mag. 18 (4) (2011) 80–92.
[6] T. Suzuki, M. Kitamura, Y. Amano, T. Hashizume, 6-DOF localization for a mobile robot using outdoor 3D voxel maps, in: Proc. IEEE/RSJ Int Intelligent Robots and Systems, IROS, Conf, 2010, pp. 5737–5743.
[7] T. Yoshida, K. Irie, E. Koyanagi, M. Tomono, A sensor platform for outdoor navigation using gyro-assisted odometry and roundly-swinging 3D laser scanner, in: Proc. IEEE/RSJ Int Intelligent Robots and Systems, IROS, Conf, 2010, pp. 1414–1420.
[8] H.E. Rauch, C.T. Striebel, F. Tung, Maximum likelihood estimates of linear dynamic systems, AIAA J. 3 (8) (1965) 1445–1450.
[9] J. Simanek, M. Reinstein, V. Kubelka, Evaluation of the EKF-based estimation architectures for data fusion in mobile robots, IEEE/ASME Trans. Mechatronics 20 (2) (2015) 985–990. http://dx.doi.org/10.1109/TMECH.2014.2311416.
[10] J. Simanek, V. Kubelka, M. Reinstein, Improving multi-modal data fusion by anomaly detection, Auton. Robots 39 (2) (2015) 139–154. URL http://dx.doi.org/10.1007/s10514-015-9431-6.
[11] M. Reinstein, M. Hoffmann, Dead reckoning in a dynamic quadruped robot based on multimodal proprioceptive sensory information, IEEE Trans. Robot. 29 (2) (2013) 563–571. http://dx.doi.org/10.1109/TRO.2012.2228309.
[12] J. Shen, D. Tick, N. Gans, Localization through fusion of discrete and continuous epipolar geometry with wheel and IMU odometry, in: Proc. American Control Conf., ACC, 2011, pp. 1292–1298.
[13] L.-P. Ellekilde, S. Huang, J.V. Miro, G. Dissanayake, Dense 3D map construction for indoor search and rescue, J. Field Robot. 24 (1–2) (2007) 71–89.
[14] P. Besl, H. McKay, A method for registration of 3-D shapes, IEEE Trans. Pattern Anal. Mach. Intell. 14 (2) (1992) 239–256.
[15] Y. Chen, G. Medioni, Object modeling by registration of multiple range images, in: Proceedings of the IEEE International Conference on Robotics and Automation, ICRA, 1991. 1991, pp. 2724–2729.
[16] A. Nuchter, K. Lingemann, J. Hertzberg, H. Surmann, 6D SLAM - 3D mapping outdoor environments, J. Field Robot. 24 (8–9) (2007) 699–722.
[17] S. Kohlbrecher, O.V. Stryk, J. Meyer, U. Klingauf, A flexible and scalable SLAM system with full 3D motion estimation, in: 2011 IEEE International Symposium on Safety, Security, and Rescue Robotics, SSRR, 2011, pp. 155–160.
[18] F. Pomerleau, F. Colas, R. Siegwart, S. Magnenat, Comparing ICP variants on real-world data sets, Auton. Robots 34 (3) (2013) 133–148.
[19] T. Oskiper, H.-P. Chiu, Z. Zhu, S. Samarasekera, R. Kumar, Multi-modal sensor fusion algorithm for ubiquitous infrastructure-free localization in vision-impaired environments, in: 2010 IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS, 2010, pp. 1513–1519. http://dx.doi.org/10.1109/IROS.2010.5649562.
[20] J. Civera, O.G. Grasa, A.J. Davison, J.M.M. Montiel, 1-Point RANSAC for extended Kalman filtering: Application to real-time structure from motion and visual odometry, J. Field Robot. 27 (5) (2010) 609–631.
[21] Y. Morales, A. Carballo, E. Takeuchi, A. Aburadani, T. Tsubouchi, Autonomous robot navigation in outdoor cluttered pedestrian walkways, J. Field Robot. 26 (8) (2009) 609–635.
[22] S.R. Sukumar, H. Bozdogan, D.L. Page, A.F. Koschan, M.A. Abidi, Sensor selection using information complexity for multi-sensor mobile robot localization, in: 2007 IEEE International Conference on Robotics and Automation, 2007, pp. 4158–4163.
[23] H. Song, V. Shin, Localization using multisensor fusion of IMM fixed lag smoother in a cricket sensor network, in: 2010 Fifth International Conference on Systems and Networks Communications, ICSNC, 2010, pp. 130–135. http://dx.doi.org/10.1109/ICSNC.2010.26.
[24] D. Lee, M. Campbell, Iterative smoothing approach using Gaussian mixture models for nonlinear estimation, in: 2012 IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS, 2012, pp. 2498–2503. http://dx.doi.org/10.1109/IROS.2012.6385752.
[25] S. Razali, K. Watanabe, S. Maeyama, K. Izumi, An unscented Rauch-Tung-Striebel smoother for SLAM problem, in: 2011 Proceedings of SICE Annual Conference, SICE, 2011, pp. 1304–1308.
[26] J. Sreenath, S. Chakrabarti, A. Sharma, Implementation of Rauch-Tung-Striebel smoother for power system dynamic state estimation in the presence of PMU measurements, in: Smart Grid Technologies-Asia (ISGT ASIA), 2015 IEEE Innovative, IEEE, 2015, pp. 1–6.
[27] S.R. McReynolds, Fixed interval smoothing-Revisited, J. Guid. Control Dyn. 13 (5) (1990) 913–921.
[28] G. Bierman, A new computationally efficient fixed-interval, discrete-time smoother, Automatica 19 (5) (1983) 503–511.
[29] K. Watanabe, S. Tzafestas, New computationally efficient formula for backward-pass fixed-interval smoother and its UD factorisation algorithm, in: Control Theory and Applications, IEE Proceedings D, Vol. 136, IET, 1989, pp. 73–78.
[30] V. Indelman, S. Williams, M. Kaess, F. Dellaert, Factor graph based incremental smoothing in inertial navigation systems, in: 2012 15th International Conference on Information Fusion (FUSION), IEEE, 2012, pp. 2154–2161.
[31] H.-P. Chiu, S. Williams, F. Dellaert, S. Samarasekera, R. Kumar, Robust vision-aided navigation using sliding-window factor graphs, in:2013 IEEE International Conference on Robotics and Automation, ICRA, 2013, pp. 46–53.
[32] M. Kaess, H. Johannsson, R. Roberts, V. Ila, J.J. Leonard, F. Dellaert, iSAM2: Incremental smoothing and mapping using the Bayes tree, Int. J. Robot. Res. (2011) 0278364911430419.
[33] S. Lange, N. Sunderhauf, P. Protzel, Incremental smoothing vs. filtering for sensor fusion on an indoor UAV, in: 2013 IEEE International Conference on Robotics and Automation (ICRA), IEEE, 2013, pp. 1773–1778.
[34] G. Sibley, L. Matthies, G. Sukhatme, Sliding window filter with application to planetary landing, J. Field Robot. 27 (5) (2010) 587–608.
[35] S. Leutenegger, S. Lynen, M. Bosse, R. Siegwart, P. Furgale, Keyframe-based visual–inertial odometry using nonlinear optimization, Int. J. Robot. Res. 34 (3) (2015) 314–334.
[36] D.G. Kottas, S.I. Roumeliotis, An iterative Kalman smoother for robust 3D localization on mobile and wearable devices, in: 2015 IEEE International Conference on Robotics and Automation (ICRA), IEEE, 2015, pp. 6336–6343.
[37] J. Ma, S. Susca, M. Bajracharya, L. Matthies, M. Malchano, D. Wooden, Robust multi-sensor, day/night 6-DOF pose estimation for a dynamic legged vehicle in GPS-denied environments, in: 2012 IEEE International Conference on Robotics and Automation, ICRA, 2012, pp. 619–626.
[38] J. Almeida, V.M. Santos, Real time egomotion of a nonholonomic vehicle using LIDAR measurements, J. Field Robot. 30 (1) (2013) 129–141.
[39] P.G. Savage, Strapdown inertial navigation integration algorithm design part 2: Velocity and position algorithms, J. Guid. Control Dyn. 21 (2) (1998) 208–221.
[40] W.G. Breckenridge, Quaternions - Proposed Standard Conventions, Tech. rep., IOM 343-79-1199, JPL, 1979.
[41] V. Kubelka, F. Pomerleau, L. Oswald, M. Reinstein, NIFTi@Zurich dataset (2013). URL https://sites.google.com/site/kubelvla/public-datasets/nifti-zurich-2013.

Vladimír Kubelka received the Ing. (M.Sc.) degree in engineering of Air and Space Systems from the Faculty of Electrical Engineering, Czech Technical University in Prague (CTU), in 2013. He is currently a Ph.D. student at the Center for Machine Perception, Dept. of Cybernetics, CTU in Prague. His research interests are self-localization and navigation of unmanned ground and air vehicles in indoor/outdoor environments for deployment in Urban Search&Rescue scenarios.

Michal Reinstein received the Ing. (M.Sc.) and Ph.D. degrees in engineering of Aircraft Information and Control Systems from the Faculty of Electrical Engineering, Czech Technical University in Prague (CTU), in 2007 and 2011, respectively. He is currently working as researcher at the Center for Machine Perception, Dept. of Cybernetics, CTU in Prague.

His research interests concern data fusion for mobile robotic systems, based on combined state estimation and machine learning techniques, intended for Urban Search & Rescue.

Tomáš Svoboda received Ph.D. in artificial intelligence and biocybernetics from the Czech Technical University in Prague in 2000 and later he spent three post-doc years with the Computer Vision Group at the ETH Zurich (Swiss Federal Institute of Technology). Currently, he is associate professor and deputy head of the department of Cybernetics at the Czech Technical University.

He has published papers on multicamera systems, omnidirectional cameras, image based retrieval, and learnable detection methods in general. His current research interests include multimodal robotic perception and object detection.

6. Tracked robot odometry for obstacle traversal in sensory deprived environment

This section presents our last paper, which is currently submitted for review in *IEEE/ASME Transactions on Mechatronics*. It is a 10-page-long paper which improves dead-reckoning localization of tracked robots in difficult terrains containing vertical obstacles.

The **introduction** explains our motivation to improve dead-reckoning localization in situations when robot passes over vertical obstacles (steps, stones, etc.). It states potential benefits brought by reducing localization drift, even for robots equipped with exteroceptive sensors. The section also lists contributions of the paper. The **related work** is presented in the following section, citing articles which deal with mechanical aspects of obstacle traversal for tracked robots, kinematic models, motion estimation and control.

Section **robot motion models** provides brief overview of dead-reckoning localization for tracked robots and shows how it fails when passing over an obstacle. New kinematic models are proposed that reduce localization drift in this scenario. A way to selecting a proper kinematic model for given situation is proposed as well.

The **experimental evaluation** shows the overall accuracy on our experimental dataset and examines selected components of the proposed track odometry. Observed results and aspects regarding the model selection are **discussed** in a separate section and finally, the **conclusion** sums up the article.

Tracked Robot Odometry for Obstacle Traversal in Sensory Deprived Environment

Vladimír Kubelka, Michal Reinstein, and Tomáš Svoboda

***Abstract*—Mobile tracked robots are suitable for traversing rough terrain. However, standard exteroceptive localization methods (visual or laser SLAM) may be unreliable due to smoke, dust, fog or insufficient lighting in harsh conditions of urban search and rescue missions. During extensive end-user evaluations in real-world conditions of such scenarios, we have observed that accuracy of dead-reckoning localization suffers while traversing vertical obstacles. We propose to combine an explicit modeling of robot kinematics and data driven approach based on machine learning. The proposed method is experimentally verified indoors and outdoors traversing various obstacles. Indoors, reference position has been recorded as well to assess accuracy of our solution. Experimental dataset is released to public to help the robotics community.**

***Index Terms*—Localization, odometry, rough terrain, tracked mobile robot.**

I. Introduction

TRACKED mobile robots are often seen in applications involving traversing uneven terrain and obstacles. Search and rescue missions after earthquakes, in underground mines or after industrial accidents are examples. Continuous self-localization while driving is necessary for robot control, sensor fusion and other functionalities.

In harsh conditions of search and rescue missions, usage of exteroceptive sensors (e.g. laser range finders, cameras) may show impossible. When testing our mobile robotic platform together with firefighters, dust, smoke and fog are often present. The robot may also enter areas with insufficient lighting conditions. During consequent exteroceptive sensors outages, standard localization approaches (e.g. visual or laser SLAM algorithms) have to be backed-up by a dead-reckoning localization which relies on proprioceptive sensors only. In tracked mobile robots, this task is performed by odometry supported by inertial measurements. It provides continuous localization but is however prone to drift (i.e. deterioration of accuracy with distance traveled) because of integration of noise in inertial sensors and track slippage.

The state-of-the art approaches to track odometry focus on correctly modeling track slippage while turning (e.g. [1], [2]). In our work, we identify another source of error – vertical obstacles – which mainly affect elevation component of position. To our best knowledge, this work is the first one that addresses problem of obstacle traversal in context of localization for tracked robots.

In order to simplify the complex problem of track-terrain interaction, we define several additional models for different stages of obstacle traversal where the robot climbs up, tips over an edge or descends down. We propose to follow a data-driven approach of training a classifier for recognizing these stages from *proprioceptive sensor* data only. Based on the classifier decision, appropriate model is chosen to govern the odometry. This allows better localization in sensory deprived environments and also provides a better starting point for exteroceptive-based localization approaches, which can later re-initialize when the robot leaves the low-visibility area.

We formulate our contributions as follows: 1) we design a dead-reckoning localization system for tracked robots which takes obstacle traversal in account, 2) we experimentally test performance of the system indoors and outdoors by experiments spanning more than 1200 m and 3) we make our dataset publicly available for robotic community in standard ROS[1] format (link given in section IV). This dataset can be useful for testing other dead-reckoning approaches for tracked robots since it contains 3D position reference for indoor experiments. The reference was acquired by using tracking capabilities of the virtual reality headset *HTC Vive*.

The paper is structured as follows: section II introduces work related to design of tracked robots for obstacle traversal and localization using track odometry. Section III describes our proposed solution, section IV summarizes experimental evaluation and section V discusses results of the experiments. Finally, section VI concludes the work.

II. Related Work

Tracked mobile robots are well suited for difficult terrains where large track-soil contact surface helps to maintain reasonable traction. However, climbing up or traversing higher obstacles is still a challenge. Many mechanical modification to the basic tracked platform concept have been proposed; work of [3] proposes a modified bogie which can operate as a normal track or rotate as a whole providing climbing ability. Alternative reconfigurable tracked robot is proposed in [4] together with stairs climbing algorithm.

Adjustable tracks – usually called *flippers* – are a common way to improve traversability, see [5], [6], [7]. Our robotic platform (Fig. 3) uses flippers as well allowing it to climb stairs and overcome vertical obstacles. There are many more possible designs that allow obstacle traversal or step climbing; authors of [8] combine walking for hard soils and tracked motion for

This work was supported by Grant Agency of the European Union under grant agreement FP7-ICT-609763 TRADR and by the Czech Technical University in Prague under grant SGS18/138/OHK3/2T/13.

V. Kubelka is with Czech Technical University in Prague, Faculty of Electrical Engineering and Czech Institute of Informatics, Robotics and Cybernetics.

M. Reinstein and T. Svoboda are with Czech Technical University in Prague, Faculty of Electrical Engineering.

[1]Robot Operating System, *http://www.ros.org*

weak soils, in [9], wheel-track robot prototype is proposed and the work of [10] describes a novel robot body configuration allowing transformation between a "claw" mode intended for climbing stairs and a "wheel" mode for flat terrain.

Approaches to adapting robot morphology to terrain are proposed in [11], [12], [13]; a computationally fast and plausible simulation of a tracked vehicle is presented in [12]. It can be utilized to train active policies which choose the best way to overcome obstacles via reinforcement learning.

In [14], authors analyze capability of a tracked robot to climb stairs. They compare different configurations of tracks and analyze climbing motion dynamics. Effect of changing payload mass held by tracked robot with flippers is investigated in [15]. The payload is a drum with an optical fiber which is being unwound as the robot proceeds forward. Our goal is to correctly estimate motion of a tracked robot during scenarios which are described in these two works.

Estimating motion of a robot based on its track velocities and other sensors is a component necessary for trajectory following. It is also useful for sensor data fusion (e.g. composing 3D lidar scans). In tracked robots, interaction between ground and tracks introduces a source of uncertainty which is difficult to model. The core of the problem is to determine amount of slippage while turning. Many approaches to solve this problem and to provide reliable odometry have been published.

Authors of [16] demonstrate that slippage of tracks can be estimated by an Extended Kalman filter by observing trajectory data of a tracked vehicle. Alternative approach to track odometry via identification of instantaneous centers of rotation is proposed by [1]. Simplified slippage model which also benefits from angular rate sensor is presented in [2], [17]; the latter also extends the odometry to 6-DOF. An approach of utilizing external position reference to identify kinematic model parameters is presented in [18]. Localization in 6-DOF by fusing track odometry and complementary-filtering-based attitude estimation is proposed in work of [19], [20], [21]. All these published approaches deal with the problem of track slippage while turning and with the associated uncertainty in velocity measurements. The extension we propose can be combined with their results by identifying moments of vertical motion and temporarily switching to our model during those instances.

Understanding kinematics of tracked vehicles is beneficial for motion control as well as demonstrated by authors of [22]. They experimentally identify kinematics of a tracked vehicle in order to design better trajectory control system. Also related to trajectory control, required torques and power consumption of tracked vehicles are investigated in [23].

To further improve motion estimation, tracks can be equipped with additional sensors which reveal points of contact with terrain. This information can be beneficial for traversing obstacles; both for robot safety (tipping over the contact point) and for localization (detecting shift of the contact point along the track). This can be done by attaching special reflectors on inner side of track segments for optical inclination measurements [24], inserting RFID technology-based sensors into track segments [25] or putting force sensors between tracks and their supports [26].

III. Robot motion models

A. Standard track odometry models

In tracked mobile robots, the main problem which complicates design of odometry systems is track slippage while turning. When a tracked robot moves forward, relative velocity between ground and tracks is zero. Yet in order to perform a turn, difference in track speeds has to be introduced and that leads to slippage. Many models for tracked robot odometry and for track-soil interaction have been proposed [16], [1], [2], [22], [23], [18].

The odometry models – in general – take track velocities as input values (v_l, v_r) and output robot velocity in the robot body frame ($\boldsymbol{v}_b = [v_x, v_y, v_z]_b^T$) and angular rate component around the z axis ω_z:

$$(\boldsymbol{v}_b, \omega_z) = f_{odom}(v_l, v_r) \tag{1}$$

Optionally, the odometry model can benefit from the angular rate ω_z being measured by an additional sensor (e.g. inertial measurement unit, IMU) [2], [17]:

$$(\boldsymbol{v}_b, \omega_z) = f_{odom,imu}(v_l, v_r, \omega_{z,imu}) \tag{2}$$

For studying the problem of climbing obstacles by a tracked mobile robot, we choose the simplest model which neglects slippage - while performing the climbing motion, assumptions about track-soil interaction are violated anyway. Nevertheless, our approach does not depend on choice of track odometry model, any of the slippage-compensating ones can be used instead. The simple model we utilize is defined as follows:

$$\boldsymbol{v}_b = \left[\frac{v_l + v_r}{2}, 0, 0\right]_b^T \tag{3}$$

$$\omega_z = \omega_{z,imu} \tag{4}$$

Knowing the robot velocity expressed in the body frame as well as the angular rate ω_z, it is possible to estimate robot position by integration of the velocity and the angular rate in time to obtain position and heading of the robot. Differential equations governing the integration process are

$$\dot{\boldsymbol{p}}_w = C^T_{(\boldsymbol{q}^b_w)} \boldsymbol{v}_b \tag{5}$$

$$\dot{\boldsymbol{q}}^b_w = \frac{1}{2} \Omega(\boldsymbol{\omega}_b) \boldsymbol{q}^b_w \tag{6}$$

where $\dot{\boldsymbol{p}}_w$ is time derivative of robot position expressed in the world coordinate frame, $C^T_{(\boldsymbol{q}^b_w)}$ is a rotation matrix transforming coordinates from robot body frame to the world frame. It is computed from a quaternion $\boldsymbol{q}^b_w$ which expresses robot attitude. (6) describes attitude update, driven by the angular rate $\boldsymbol{\omega}_b = [0, 0, \omega_z]^T$. $\Omega(\omega)$ stands for

$$\Omega(\omega) = \begin{bmatrix} 0 & \omega_3 & -\omega_2 & \omega_1 \\ -\omega_3 & 0 & \omega_1 & \omega_2 \\ \omega_2 & -\omega_1 & 0 & \omega_3 \\ -\omega_1 & -\omega_2 & -\omega_3 & 0 \end{bmatrix} \tag{7}$$

according to [27, eq. 108]. Note that this is the most basic data integration scheme, which can be replaced be an advanced data fusion filter incorporating other external measurements [18], [20], [21] or the velocity measurements can be used in a factor graph fusion approach.

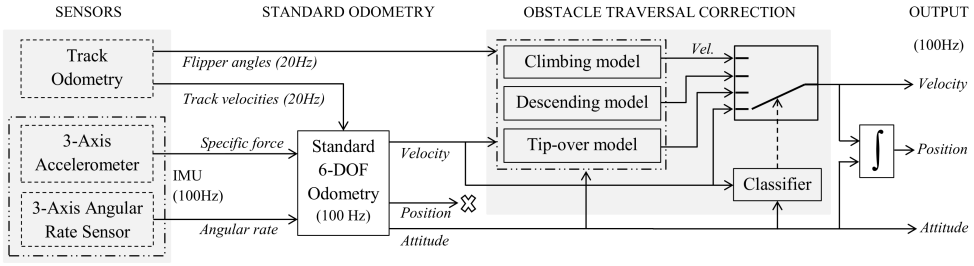


Fig. 1. Block diagram of the proposed track odometry system. Standard 6-DOF odometry with input from proprioceptive sensors computes intermediate velocity, position and robot attitude. These values are corrected in the OTC block based on classifier decision to either keep the standard output or to switch to one of the proposed obstacle traversing modes.

The odometry models mentioned so far consider only planar motion, extension to 3D is described in [17], [20]. In 3D, attitude is obtained from inertial measurements and corrected by magnetometer data if available. Eq. 5 still holds, the only difference is that the quaternion $\boldsymbol{q}_w^b$ expresses non-zero roll and pitch angles as well. Our approach requires this extension to 3D since we focus on describing motion of traversing vertical obstacles.

B. Obstacle traversal problem

Standard approaches to track odometry described in section III-A provide good results as long as all contact points with terrain lie on the flat part of the track. Fig. 2 and 3 demonstrate a case when this condition does not hold, a track bogie of the robot faces a step.

As soon as the track begins to move forward, physical interaction of the front round section of the track with the obstacle forces the bogie to climb upwards [14, Fig.3], see the top part of Fig. 3. However, standard approaches to track odometry indicate forward motion based on measured track velocities (3D odometry will eventually estimate position increment in the z axis due to rising pitch angle). Estimated trajectory is shown in the bottom part of the Fig. 3.

Similar problem arises from tipping over an edge (as shown in Fig. 5) or descending down from an obstacle. As the robot traverses the obstacle, position of contact points between tracks and terrain changes and so does the odometry model. Our solution to this problem lies in designing separate models for different stages of the traversal and finally training a classifier that automatically selects one of these models based on available sensor data.

Fig. 1 provides an overview of our proposed solution. Raw sensor data are processed by a standard odometry system. It estimates velocity expressed in the body frame of the robot, position of the robot – which is however discarded – and attitude of the robot (following (3), (5) and (6) respectively).

In the **O**bstacle **T**raversal **C**orrection section (OTC), velocity estimated by the standard odometry is corrected by one of **O**bstacle **T**raversal **M**odels (OTM) or left untouched for flat terrains. The decision which model is appropriate is made by the classifier block. Following (5), new corrected position is computed and output together with attitude estimates, which is not modified in our approach. Following subsections describe the proposed components in detail.

Fig. 2. Example of a vertical obstacle that poses a problem for track odometry.

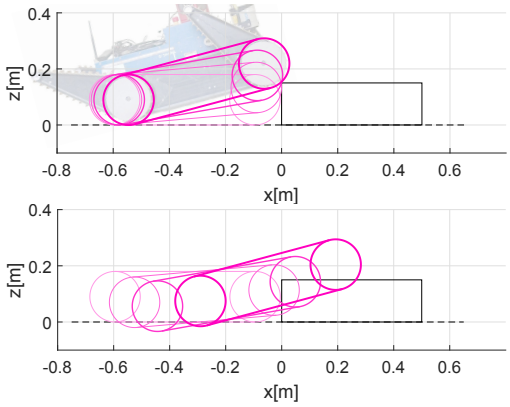


Fig. 3. Climbing a step with a single-track robot. Real motion (top) and trajectory estimated by a standard 3D odometry algorithm (bottom).

C. Obstacle climbing and descending models

Climbing obstacles poses a problem for standard track odometry. We model it by the OTM that takes pitch angle of the robot as input and computes position of the robot relative to the obstacle. It is intended for tracked robots with or without active sub-tracks (often called *flippers*). The model assumes climbing symmetrical obstacles (however, we evaluate its performance on non-symmetrical obstacles as well).

Fig. 4 demonstrates the core idea of the model. It consists of two line segments representing bottom sides of the track assembly (main track and a flipper, painted in violet) and the obstacle (black rectangle).

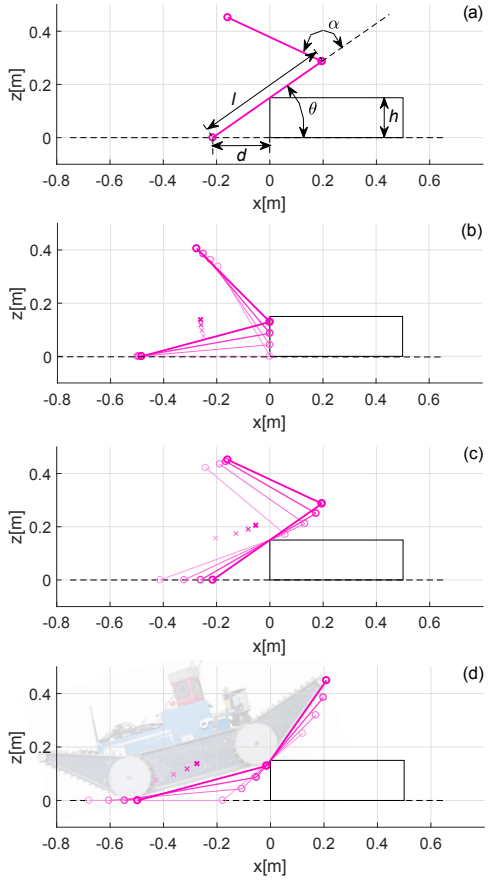


Fig. 4. Kinematic model for obstacle climbing. From top: (a) model parameters, (b,c) climbing with flippers retracted, (d) climbing using flippers. The violet x symbol denotes motion of the origin of the robot frame.

Input values are pitch angle of the main track θ, flipper angle relative to the main track α, obstacle height[2] h and track length l. The model outputs distance d between the obstacle and the point the track touches the ground, measured on the x axis. It is computed as follows:

$$d = \frac{h\cos(\theta+\alpha) + l\cos(\theta)\sin(\theta+\alpha) - l\sin(\theta)\cos(\theta+\alpha)}{\sin(\theta+\alpha)} \tag{8}$$

See the plot (a) in Fig. 4 which shows all the values in the geometry of the model.

Transition between the obstacle contact point lying on the flipper (Fig. 4, d) or on the main track (Fig. 4, c) is triggered by this condition:

$$|l\sin(\theta)| \geq h \tag{9}$$

If this inequality holds, the contact point lies on the main track bogie. By substituting $\alpha = 0$ in (8), we adjust the model to search for the contact point on the main track.

[2]Without additional sensors, obstacle height cannot be measured directly. If such measurements are not available, we propose a sufficiently robust approach in section III-F.

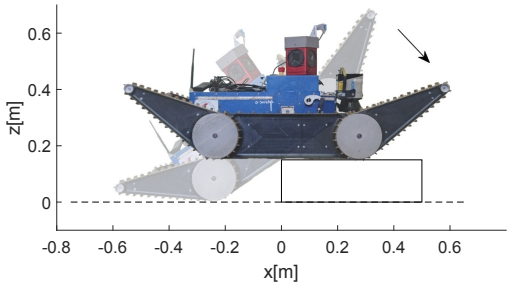


Fig. 5. Tipping over the obstacle edge after climbing it up.

The input angles α and θ cannot be arbitrary. If the flipper angle relative to horizontal plane $\theta + \alpha$ exceeds 90 degrees, it cannot be interacting with an obstacle anymore – at least according to our model. In that case, we decrease value of α until this condition holds:

$$(\theta + \alpha) = \frac{\pi}{2} \tag{10}$$

It leads to a motion estimate which is equivalent to climbing up an obstacle using only the main track bogie (as shown in the plot (b) of Fig. 4). This is also the way to model obstacle climbing for robots without front flippers.

The model checks for impossible configurations as well for which

$$((\theta + \alpha) \leq 0) \vee (d > (l + \text{flipper length})) \tag{11}$$

becomes true. This condition indicates either a situation when the robot "*dives*" into ground below the obstacle or the d distance is greater than the actual reach of flippers.

Descending obstacles can be viewed as climbing backwards; geometry of the problem differs only in the sign of the pitch angle θ and in the sign of the d distance. Moreover, the α angle refers to the angle of rear flippers. After substituting into (8), we obtain an OTM for obstacle descending.

To compute velocity measurement from the models, two distances d_i and d_{i+1} are computed from two subsequent pitch values (on our platform, pitch angle is sampled at 100 Hz). Since there is a rigid transformation between the robot frame origin and the main track, d_i and d_{i+1} together with respective robot pitch angles are sufficient to compute robot frame origin motion (denoted by the violet x symbols in Fig. 4) relative to the obstacle. This motion is expressed in the body frame of the robot as an approximation of velocity with non-zero v_x and v_z components (in standard odometry, only v_x is non-zero).

D. Tip-over motion model

Another source of localization error is tip-over motion on edges of the obstacle (see Fig. 5). In this situation, the robot tips over by its own weight and motion of its tracks is usually minimal – only to push the robot over the tipping point. However, translation of the robot origin relative to the world causes position estimation error if not treated correctly.

We design the OTM to correct this error by estimating translation of the origin of the robot frame expressed in the

robot frame. This is obtained by rotating the robot frame origin around the tipping point from zero pitch angle to two subsequent measured pitch angles:

$$\boldsymbol{x}_{i,w'} = -C^T_{(\theta_i)}\boldsymbol{x}_{\text{t.p.}} \tag{12}$$

$$\boldsymbol{x}_{i+1,w'} = -C^T_{(\theta_{i+1})}\boldsymbol{x}_{\text{t.p.}} \tag{13}$$

where $\boldsymbol{x}_{\text{t.p.}}$ are coordinates of the tipping point in the x, z plane ($y = 0$) and $C^T_{(\theta_i)}$ rotation matrix around y axis computed from pitch angle θ_i. The results $\boldsymbol{x}_{i,w'}$ and $\boldsymbol{x}_{i+1,w'}$ are expressed in a temporary world frame denoted by w' which is aligned with the robot body frame in position and yaw angle. In this frame, we compute position increment and express it back in the robot body frame:

$$\Delta\boldsymbol{x}_b = C_{(\theta_i)}\left(\boldsymbol{x}_{i+1,w'} - \boldsymbol{x}_{i,w'}\right) \tag{14}$$

This increment is an approximation of a velocity measurement with non-zero v_x and v_z components as well.

E. Switching obstacle traversal models

Each OTM works best only in situation it is designed for. It is therefore important to correctly switch between them as the robot traverses an obstacle. For example, OTM for obstacle climbing utilized while traversing flat terrain will probably trigger the fail-safe condition (11).

To make the OTM switching automatic, we propose to train a classifier (e.g. SVM, [28], [29]) based on measured dynamics of the robot. The 4-class classifier decides which odometry model is appropriate for the given moment. Feature vector consists of three components

$$\boldsymbol{f} = [\frac{v_l + v_r}{2}, \dot{\theta}, \theta] \tag{15}$$

where $\frac{v_l+v_r}{2}$ is forward speed measured by tracks and $\dot{\theta}$ is pitch angular rate. The classifier is trained in a supervised manner by providing labels for a representative set of training experiments, which contain obstacles, flat terrain and inclined ramps. To mark the stage labels manually, we use our reference tracking system described in the section IV. The procedure involves selecting intervals, where one of the tree OTMs should be used for predicting motion. The rest is labeled as standard odometry. Distribution of labels in the feature space is shown in the Fig. 6. The classes are well separated, thus we choose a fast linear classier – SVM. The SVM is learned utilizing the *LIBSVM* [29] implementation, with a multi-class classifier and radial-basis kernel. Required performance of the classifier is discussed in the section IV-C.

F. Velocity components selection

One of the model inputs is obstacle height h. Its value – paradoxically – does not affect estimated position increments in the vertical z axis. While climbing up an obstacle or descending one, accumulated height depends on track lengths and pitch differences. However, obstacle height affects x component of position increments.

Without external or additional proprioceptive sensors, estimating obstacle height h has shown to be difficult. Experiments suggest – however – that the obstacle height can be approximated by a constant value equal to half of maximum possible height the robot can climb. We verify that this approximation is acceptable in section IV. Directly measuring h may by possible by adding additional tactile sensors [24], [25], [26]. Nevertheless, we aim for a solution which works for general tracked robots with minimum number od sensors and which is sufficiently robust even in harsh conditions of sensory deprived environment.

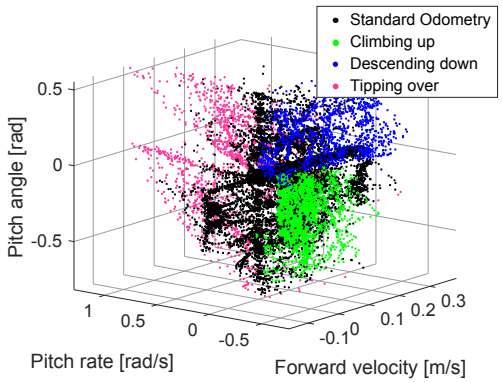


Fig. 6. Learning data samples for the SVM classifier. Labels created manually using the HTC Vive reference system.

Due to the fact our OTMs are still an approximation of the real world physical interaction, value of the v_x velocity component computed by the model is slightly underestimated (distribution of the error fits a Gaussian with negative mean on our dataset). We propose a hyperparameter-free solution and a tunable solution to this problem and compare them in the section IV.

The hyperparameter-free solution utilizes only the v_y and v_z velocity components estimated by OTM and keeps the v_x component reported by the standard odometry.

The tunable approach originates from the observation on experiments with positional reference: standard odometry usually over-estimates the v_x component (error can be approximated by a Gaussian with positive mean). True value of v_x is therefore expected to lie between values estimated by standard odometry and OTMs. To fine-tune the model to reflect the reality, we propose to combine them:

$$\boldsymbol{v}_b = [wv_{x,o} + (1-w)v_{x,k}, v_{y,k}, v_{z,k}]^T_b \tag{16}$$

where $\boldsymbol{v}_b$ is velocity expressed in the body frame, $w \in <0, 1>$ is a weighting factor, subscript o denotes the original odometry component and the k subscript denotes our OTM-based components. The first solution is equivalent to $w = 1$; the optimal value of the weight can be found by minimizing localization error over testing set of experiments. On flat terrains however, only the standard track odometry $\boldsymbol{v}_{b,o}$ is used.

IV. Experimental evaluation

For position reference, we use a virtual reality headset *HTC Vive*[3]. Using driver for *Robotic Operating System* (ROS) [30],

[3] *https://www.vive.com*

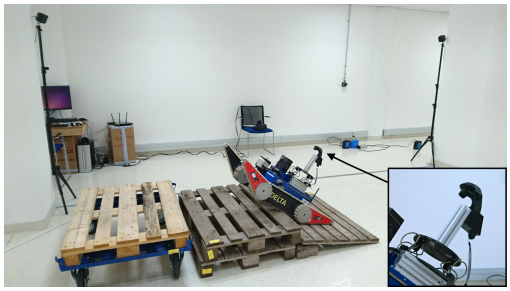


Fig. 7. Indoor experiment with HTC Vive reference system. Detail shows a HTC Vive controller attached to the robot via vibration dampener, which is necessary due to an IMU inside the tracked controller.

we obtain coordinates of a wireless controller which comes with the headset system. When properly attached to the robotic platform (see Fig. 7), calibration procedure allows us to find transformation from a reference frame of the controller to body frame of the robot. Accuracy of the reference system is $\pm 10\,\mathrm{mm}$.

The robotic platform used for experiments (see Fig. 2) is a mobile robot developed within the *TRADR*[4] project. Locomotion system consists of two main tracks, each of them is equipped with two flippers. For attitude estimation, an Xsens MTi-G IMU is mounted near the robot frame origin.

We have performed both indoor and outdoor experiments. Indoors, we have selected a representative set of experiments to train the SVM classifier (16 experiment runs, $299\,m$ total distance). The rest of the indoor experiments serve as a testing set (22 runs, $457\,m$). Some of these experiments miss the HTC Vive reference (because of technical limitations of the reference system, e.g. experiments through staircases); we provide manual measurements (e.g. staircase height) instead.

Outdoor experiments (8 runs, $455\,\mathrm{m}$) were performed on a long stairway (see Fig. 13) and inside an industrial safety training plant[5]. Because of GPS signal attenuation through trees in the case of the stairway or reflections from nearby metal construction at the training plant, standard deviation of indicated GPS position is in the order of meters. Since the localization errors we aim to correct are in magnitude of decimeters, such reference system cannot be used. Instead, we plan trajectories of the experiments to have identical start and end points. This way, final localization error can be measured.

All experiments were released to public in the form of ROS compatible datasets (including the ground truth)[6]. They are stored as ROS *Indigo* bag files and thus can be re-played and viewed by the *Rviz* tool.

[4]TRADR: Long-Term Human-Robot Teaming for Robot-Assisted Disaster Response *http://www.tradr-project.eu*

[5]RDM Training Plant *http://www.rdmtrainingplant.nl*

[6]Dataset at *https://sites.google.com/site/kubelvla/public-datasets/obstacle-traversal-dataset-2017* See *http://ros.org* for the file format information

A. Performance metrics

To evaluate the experiments, we prefer *Average Position Error* (APE) metric:

$$\mathrm{APE}(t) = \frac{\sum_{i=1}^{k} \sqrt{(\hat{x}_i - x_i)^2 + (\hat{y}_i - y_i)^2 + (\hat{z}_i - z_i)^2}}{k(t)} \tag{17}$$

where x, y and z are estimated position coordinates, $k(t)$ mapping from time t to discrete time k and $\hat{x}_i$, $\hat{y}_i$ and $\hat{z}_i$ is ground truth position. To focus on effects on a selected axis, we modify the APE into *Average Component Error* (ACE) as follows:

$$\mathrm{ACE}_z(t) = \frac{\sum_{i=1}^{k} |\hat{z}_i - z_i|}{k(t)} \tag{18}$$

where the subscript z denotes the axis of interest. APE and ACE are well suited to show the effects of our proposed odometry (especially drift in the z axis) for trajectories with varying length.

For experiments without full reference, we evaluate difference between the starting point and the final point of the robot trajectory – they are supposed to be identical by design of the experiments.

B. Training classifier for Obstacle Traversal Correction model switching

The first subset of experiments serves as a training set for the SVM. In 16 runs spanning $299\,m$ traveled, robot passes obstacles of various heights made of pallets, inclined ramps and flat surfaces.

Example of typical performance of our proposed OTC for these experiments is demonstrated in Fig. 8. True and estimated trajectories[7] of the robot passing over pallets are shown in Fig. 8. Two pallet heights h and three different front flipper angle α settings are shown. In the third run (bottom plot), front flippers ($\alpha = 100°$) do not contribute to the motion of the robot during the experiment.

C. Effect of the classifier performance to the odometry accuracy

Fig. 9 demonstrates effect of false classifications on the estimated trajectory. For a selected testing experiment (similar to the one depicted in Fig. 10), appropriate OTMs are chosen manually i.e. the same way we create the ground truth for training. Yet a given number of these manual classifications are perturbed to a different class.

The perturbation is done as follows: for given accuracy (percentage of correctly predicted labels over all predictions), random subset of labels is selected (e.g. for accuracy 60%, the randomly chosen subset contains 40% of all labels). Each label in this subset is changed to a new random label, different from the original.

We run the odometry corrected by OTC and store its localization accuracy. To obtain different subsets of these false classifications, the process is repeated twenty times for each

[7]We show only side view since the robot moves straight forward and the y coordinate changes only a little

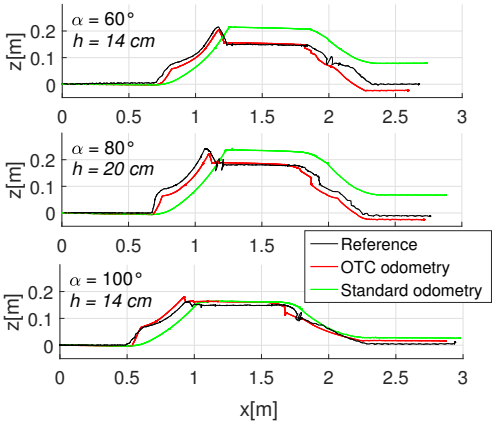


Fig. 8. Example of three passes over wooden pallets 14 cm and 20 cm high. The indicated α angle is the flipper angle while climbing up the pallet, h stands for height of the pallet.

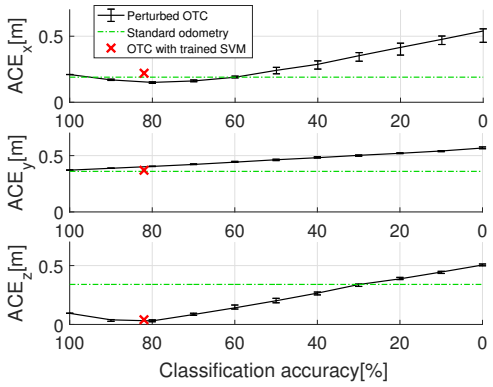


Fig. 9. Effect of label perturbation on localization accuracy when applying OTC. Results of standard odometry (black solid line) and OTC governed by trained classifier shown for comparison. Error bars denote lower and upper quartiles.

TABLE I
CLASSIFIER CONFUSION MATRIX – TRAINING

	Predicted labels			
True labels	Standard odometry	Climbing up	Descending down	Tipping over
Standard odometry	**24616**	2044	1001	1019
Climbing up	1688	**16595**	0	1
Descending down	1007	6	**9548**	2
Tipping over	2595	62	49	**5732**

TABLE II
CLASSIFIER CONFUSION MATRIX – TESTING

	Predicted labels			
True labels	Standard odometry	Climbing up	Descending down	Tipping over
Standard odometry	**51231**	499	273	296
Climbing up	3763	**696**	0	13
Descending down	2739	79	**1277**	8
Tipping over	4122	3	8	**290**

classification accuracy. Resulting ACE is shown as error bars in Fig. 9.

Compared to standard odometry, classification accuracy better then 60% improves overall localization accuracy. Results show drop of localization accuracy in the interval between 100% and 80% classifier accuracy. Cause of this drop is manually created ground truth which cannot be perfect – random perturbation can actually improve the final result. Also note that the x and y ACE is roughly the same for the OTC approach (provided high classifier accuracy) and for the standard odometry. This is caused by the drift of the yaw angle being the main contributor to the x and y position error. Attitude is not corrected in our approach and thus, these errors remain approximately the same.

Tables I and II contain confusion matrices of our trained SVM classifier. Classification of testing data (selected representative long experiments not included into training data) shows high rate of mis-classifications into the first class. Overall classification accuracy of the trained SVM is 97% and 82% on the training and testing data respectively. Effect of mis-classification to localization accuracy is discussed in section V.

D. Evaluation on testing dataset with 3D reference

This section presents the overall evaluation of the whole chain of standard odometry, OTC and trained classifier. Subset of indoor experiments with 3D reference, disjoint with the training subset, is used to study localization error and to build overall accuracy statistics. This testing subset consists of 16 runs spanning 288 m. Passing over obstacle in non-zero roll angles as well as various inclined ramps is included to test the odometry behavior in unforeseen conditions[8].

Table III sums up overall statistics of the testing set of experiments. From each experiment, final value of ACE and APE is taken, i.e. t = experiment length. The standard odometry results are the baseline. When applying OTC with the hyperparameter-free setting (weight $w = 1$ in (16)), trajectories almost overlap when projected into the x, y subspace. The z axis drift is however reduced to half of the original. The result can be improved further by optimizing over the weight w when the tunable approach is chosen. We present $w = 0.5$ setting which we experimentally verified works the best for minimizing the z axis drift. With this weight, x, y subspace gains some improvement in the x coordinate and looses in the y coordinate. Yet comparing the lower and upper quartiles, the error distributions overlap and we consider this difference negligible. Following experiments in this section use $w = 0.5$ weight.

Example of the testing experiments is shown in Fig. 10 and Fig. 11: it involves traversing three pallets in circles. Localization error in x and y coordinates grows as the yaw angle drifts. OTC module does not modify estimated attitude

[8]On our public dataset website, see *Test indoor experiments*, *Three steps structure* and *Long trajectories over pallets, indoors* for additional photos and details.

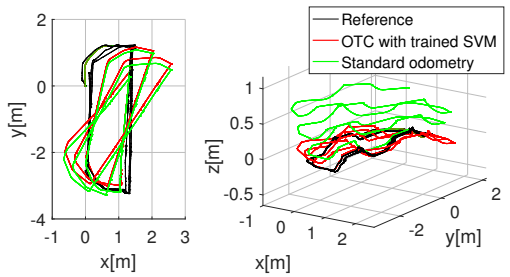


Fig. 10. Top and side views of a trajectory obtained from an example testing experiment. In this experiment, the robot navigates over three pallets in circle, several times.

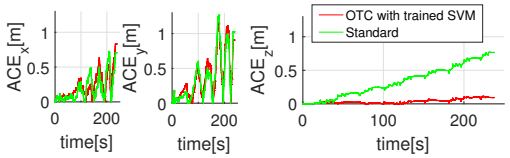


Fig. 11. Average component error for trajectory shown in Fig. 10

of the robot and thus, this result follows result of standard odometry. OTC – however – decreases drift in the z axis which is to be expected as we try to improve modeling of the obstacle climbing and descending.

E. Outdoor and staircase testing experiments

Series of outdoor experiments, which are part of the public dataset, was recorded at the industrial safety training plant (6 runs, 145 m). During training session with firefighters (end-users in the TRADR project), harsh conditions of search and rescue missions were simulated. The robot was teleoperated through scenario involving climbing and descending obstacles present in the area. Due to lack of reference system, start and end point of the trajectories were identical – total drift of the robot position could be measured this way. One of the runs is shown in Fig. 12. Two times, the robot climbs onto a concrete platform (first time over two pallets, then directly over the concrete step). The z coordinate plot shows reduction of drift in this axis. Following trend from indoor testing experiments, x and y position components are affected only slightly. However, the z component drift is reduced, its median final error over six experiments relative to standard odometry is 43% (i.e. 57% improvement). Final error depends

TABLE III
FINAL ACE AND APE OF TESTING EXPERIMENTS

	Standard odometry	OTC with trained SVM	
		$w = 1$	$w = 0.5$
ACE_x	0.14|**0.20**|0.25	0.13|**0.20**|0.25	0.13|**0.19**|0.23
ACE_y	0.11|**0.17**|0.22	0.12|**0.17**|0.23	0.13|**0.20**|0.25
ACE_z	0.053|**0.093**|0.16	0.024|**0.048**|0.091	0.019|**0.034**|0.056
APE	0.25|**0.31**|0.39	0.24|**0.31**|0.36	0.23|**0.30**|0.35

lower|**middle**|upper quartile [m]

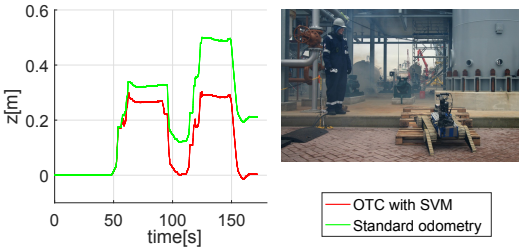


Fig. 12. Experiment performed in the industrial training plant. The robot climbs up and descends down from an elevated concrete platform. For testing purpose, this is repeated twice – once it climbs up using the pile of pallets, the second time, it climbs the concrete step itself. Plot of the z component of the estimated trajectory shows drift of the standard odometry.

Fig. 13. Climbing an 70-meters-long exterior staircase with our robotic platform. It was additionally equipped with a robotic arm (right image) to test different weight distribution effect on the trajectory estimate and model robustness.

on number of obstacles traversed by the robot; difficult terrain yields more improvement.

In order to test our approach on another typical vertical obstacle, we climbed and descended several staircases, indoor (6 runs, 157 m) and outdoor (2 runs, 310 m). Fig. 13 shows the robot climbing up the outdoor stairway (with steps far away from each other and cca. 16 m elevation gain). Median value of final error relative to standard odometry is 73% for all staircase experiments (27% improvement).

To push the algorithm to the limits and to test the robustness of our proposed models, we mounted a robotic arm on the robot to alter its weight distribution during one of the outdoor stairway experiments. The extra weight causes change of center of mass of the robot and thus different forces acting on flippers while climbing the stairs. However, we observed no major change in localization accuracy compared to runs without the extra weight.

V. DISCUSSION

Main advantage of the proposed odometry system, observed in all experiments, is reduction of drift in the z (elevation) axis resulting in smaller final localization error ACE_z. This drift occurs mostly when traversing vertical obstacles (as demonstrated in Fig. 8) and thus, improvement over standard odometry can be observed in experiments that contain such obstacles. The other two components (x, y) are mainly affected by yaw angle drift which has to be addressed by other means (better IMU calibration, fusion with magnetometer data if possible etc.). Nevertheless, OTC also corrects x increments when overcoming obstacles, as shown in Fig. 8.

Even with high rate of mis-classifications observed in Table II, our solution improves accuracy of localization. Most

of the mis-classifications are into standard odometry class and the error increments only as much as it would without OTC. In the case of mis-classification into one of the remaining classes, OTC still checks for impossible configurations and can refuse incorrect classifier decisions.

VI. Conclusion

We propose a novel approach to 3D odometry for tracked robots which focuses on the problem of traversing vertical obstacles. This approach relies on standard proprioceptive sensors only and can be deployed in low-visibility conditions of sensory-deprived environments, often seen in search and rescue missions. We extend standard odometry by additional kinematic models (climbing, tipping-over and descending) which all together cover traversing vertical obstacles and flat terrains. The problem of selecting appropriate models for flat terrain or different stages of obstacle traversal is solved by an SVM classifier trained on proprioceptive measurements. This way we effectively combine explicit kinematics modeling with data driven approach using machine learning.

Proposed approach is experimentally verified by indoor and outdoor experiments, which we release to the public to help and challenge the robotics community. For indoor experiments, we provide accurate position reference which allows comparison between estimated and true trajectories. Outdoor experiments include trajectories from an industrial training plant recorded during a search and rescue exercise session with a firefighter brigade.

While traversing obstacles, more than 50% improvement is observed in the z-axis position error (ACE_z). Our solution can be applied to robots with or without front flippers; we also demonstrate that changes in weight distribution of the robot do not affect localization accuracy. In our opinion, decreased position error drift of a dead-reckoning localization system is beneficial in environments with obstacles and low-visibility conditions which can impede visual or laser localization.

References

[1] J. L. Martnez, A. Mandow, J. Morales, S. Pedraza, and A. Garca-Cerezo, "Approximating kinematics for tracked mobile robots," *The International Journal of Robotics Research*, vol. 24, no. 10, pp. 867–878, 2005. [Online]. Available: http://dx.doi.org/10.1177/0278364905058239

[2] D. Endo, Y. Okada, K. Nagatani, and K. Yoshida, "Path following control for tracked vehicles based on slip-compensating odometry," in *2007 IEEE/RSJ International Conference on Intelligent Robots and Systems*, Oct 2007, pp. 2871–2876.

[3] G. Lan and S. Ma, "Step-climbing analysis of a novel tracked robot," in *2005 IEEE International Conference on Robotics and Biomimetics - ROBIO*, 2005, pp. 544–549.

[4] P. Ben-Tzvi, S. Ito, and A. A. Goldenberg, "Autonomous stair climbing with reconfigurable tracked mobile robot," in *2007 International Workshop on Robotic and Sensors Environments*, Oct 2007, pp. 1–6.

[5] W. Wang, Z. Du, and L. Sun, "Dynamic load effect on tracked robot obstacle performance," in *2007 IEEE International Conference on Mechatronics*, May 2007, pp. 1–6.

[6] T. Fujita and T. Shoji, "Development of a rough terrain mobile robot with multistage tracks," in *2013 16th International Conference on Advanced Robotics (ICAR)*, Nov 2013, pp. 1–6.

[7] S. Singh, B. D. Jadhav, and K. M. Krishna, "Posture control of a three-segmented tracked robot with torque minimization during step climbing," in *2014 IEEE International Conference on Robotics and Automation (ICRA)*, May 2014, pp. 4200–4207.

[8] K. Nagatani, H. Kinoshita, K. Yoshida, K. Tadakuma, and E. Koyanagi, "Development of leg-track hybrid locomotion to traverse loose slopes and irregular terrain," *Journal of Field Robotics*, vol. 28, no. 6, pp. 950–960, 2011. [Online]. Available: http://dx.doi.org/10.1002/rob.20415

[9] W. Guo, Y. Mu, and X. Gao, "Step-climbing ability research of a small scout wheel-track robot platform," in *2015 IEEE International Conference on Robotics and Biomimetics (ROBIO)*, Dec 2015, pp. 2097–2102.

[10] L. H. Pan, C. N. Kuo, C. Y. Huang, and J. J. Chou, "The claw-wheel transformable hybrid robot with reliable stair climbing and high maneuverability," in *2016 IEEE International Conference on Automation Science and Engineering (CASE)*, Aug 2016, pp. 233–238.

[11] K. Zimmermann, P. Zuzánek, M. Reinstein, T. Petříček, and V. Hlaváč, "Adaptive traversability of partially occluded obstacles," in *2015 IEEE International Conference on Robotics and Automation (ICRA)*, May 2015, pp. 3959–3964.

[12] M. Pecka, V. Šalanský, K. Zimmermann, and T. Svoboda, "Autonomous flipper control with safety constraints," in *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Oct 2016, pp. 2889–2894.

[13] M. Gianni, M. A. R. Garcia, F. Ferri, and F. Pirri, "Terrain contact modeling and classification for atvs," in *2016 IEEE International Conference on Robotics and Automation (ICRA)*, May 2016, pp. 186–192.

[14] J. Liu, Y. Wang, S. Ma, and B. Li, "Analysis of stairs-climbing ability for a tracked reconfigurable modular robot," in *IEEE International Safety, Security and Rescue Rototics, Workshop, 2005.*, June 2005, pp. 36–41.

[15] Y. Li, S. Ge, H. Fang, C. Chu, and Y. Liu, "Effects of the fiber releasing on step-climbing performance of the articulated tracks robots," in *2009 IEEE International Conference on Robotics and Biomimetics (ROBIO)*, Dec 2009, pp. 818–823.

[16] A. T. Le, D. C. Rye, and H. F. Durrant-Whyte, "Estimation of track-soil interactions for autonomous tracked vehicles," in *Proceedings of International Conference on Robotics and Automation*, vol. 2, Apr 1997, pp. 1388–1393 vol.2.

[17] K. Nagatani, N. Tokunaga, Y. Okada, and K. Yoshida, "Continuous acquisition of three-dimensional environment information for tracked vehicles on uneven terrain," in *2008 IEEE International Workshop on Safety, Security and Rescue Robotics*, Oct 2008, pp. 25–30.

[18] J. Pentzer, S. Brennan, and K. Reichard, "Model-based prediction of skid-steer robot kinematics using online estimation of track instantaneous centers of rotation," *Journal of Field Robotics*, vol. 31, no. 3, pp. 455–476, 2014. [Online]. Available: http://dx.doi.org/10.1002/rob.21509

[19] M. Reinstein and M. Hoffmann, "Dead Reckoning in a Dynamic Quadruped Robot Based on Multimodal Proprioceptive Sensory Information," *IEEE Transactions on Robotics*, vol. 29, no. 2, pp. 563–571, April 2013.

[20] J. Simanek, M. Reinstein, and V. Kubelka, "Evaluation of the ekf-based estimation architectures for data fusion in mobile robots," *IEEE/ASME Transactions on Mechatronics*, vol. 20, no. 2, pp. 985–990, April 2015.

[21] V. Kubelka, L. Oswald, F. Pomerleau, F. Colas, T. Svoboda, and M. Reinstein, "Robust data fusion of multimodal sensory information for mobile robots," *Journal of Field Robotics*, vol. 32, no. 4, pp. 447–473, 2015. [Online]. Available: http://dx.doi.org/10.1002/rob.21535

[22] S. A. A. Moosavian and A. Kalantari, "Experimental slip estimation for exact kinematics modeling and control of a tracked mobile robot," in *2008 IEEE/RSJ International Conference on Intelligent Robots and Systems*, Sept 2008, pp. 95–100.

[23] T. Guo and H. Peng, "A simplified skid-steering model for torque and power analysis of tracked small unmanned ground vehicles," in *2013 American Control Conference*, June 2013, pp. 1106–1111.

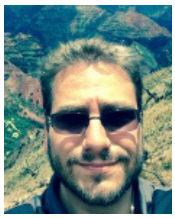
[24] D. Inoue, M. Konyo, K. Ohno, and S. Tadokoro, "Contact points detection for tracked mobile robots using inclination of track chains," in *2008 IEEE/ASME International Conference on Advanced Intelligent Mechatronics*, July 2008, pp. 194–199.

[25] L. Heckes, T. Predki, and P. Labenda, "A wireless contact point sensor for tracked robots," in *2012 IEEE International Conference on Technologies for Practical Robot Applications (TePRA)*, April 2012, pp. 86–90.

[26] V. Šalanský, V. Kubelka, K. Zimmermann, M. Reinstein, and T. Svoboda, "Touching without vision: terrain perception in sensory deprived environments," in *Proceedings of the 21st Computer Vision Winter Workshop*, L. Čehovin, R. Mandeljc, and V. Štruc, Eds., Slovenian Pattern Recognition Society. Ljubljana, Slovenia: Slovenian Pattern Recognition Society, February 2016, p. 9.

[27] N. Trawny and S. I. Roumeliotis, "Indirect Kalman Filter for 3D Attitude Estimation - A Tutorial for Quaternion Algebra," University of Minnesota, Tech. Rep., 2005.
[28] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, no. 3, pp. 273–297, Sep 1995. [Online]. Available: https://doi.org/10.1007/BF00994018
[29] C.-C. Chang and C.-J. Lin, "LIBSVM: A library for support vector machines," *ACM Transactions on Intelligent Systems and Technology*, vol. 2, pp. 27:1–27:27, 2011, software available at http://www.csie.ntu.edu.tw/∼cjlin/libsvm.
[30] RoboSavvy, "vive_ros," https://github.com/robosavvy/vive_ros, 2016.

Vladimír Kubelka received the Ing. (M.Sc.) degree in engineering of Air and Space Systems from the Faculty of Electrical Engineering, Czech Technical University in Prague (CTU), in 2013. He is currently a Ph.D. student at the Center for Machine Perception, Dept. of Cybernetics, CTU in Prague. His research interests are self-localization and navigation of unmanned ground and air vehicles in indoor/outdoor environments for deployment in Urban Search & Rescue scenarios.



Michal Reinstein received the Ing. (M.Sc.) and Ph.D. degrees in engineering of Aircraft Information and Control Systems from the Faculty of Electrical Engineering, Czech Technical University in Prague (CTU), in 2007 and 2011, respectively. He is currently working as researcher at the Center for Machine Perception, Dept. of Cybernetics, CTU in Prague.

His research interests concern data fusion for mobile robotic systems, based on combined state estimation and machine learning techniques, intended for Urban Search & Rescue.

Tomáš Svoboda received Ph.D. in artificial intelligence and biocybernetics from the Czech Technical University in Prague in 2000 and later he spent three post-doc years with the Computer Vision Group at the ETH Zurich (Swiss Federal Institute of Technology). Currently, he is associate professor and the chair of department of Cybernetics at the Czech Technical University.

He has published papers on multicamera systems, omnidirectional cameras, image based retrieval, and learnable detection methods in general. His current research interests include multimodal active robotic perception and exploration.

7. Discussion

In this section, we address topics related to robot localization and advances in the field, which appeared after our publications. We mainly focus on publications from years 2015-2018. We also provide a brief overview of techniques related to the localization for search and rescue, which we have not addressed in our publications, and discuss their relation to our work.

7.1. Magnetic field measurements

In low-visibility environments, exploiting Earth magnetic field offers additional means to estimate heading or velocity of the robot. Sufficiently far from structures affecting Earth magnetic field, utilizing magnetometer as a compass is a popular way to estimate the absolute robot heading. However, the attention must be paid to shielding the sensor from additive sources of magnetic field and from ferromagnetic materials generally. It is possible to comply with these requirements in the aerospace industry and in drones, which are made of non-ferromagnetic materials.

Ground robots, however, often operate in environments where magnetic disturbances cause heading estimate errors. On the other hand, the magnetic field perturbed by structures present in buildings can be utilized to estimate robot velocity by measuring changes in gradient of the field; experiments performed by [4, 5, 6] show that dead-reckoning localization based on magnetometer and IMU is possible. Yet, problems caused by magnetic field disturbances introduced by the robot itself and by AC current in the electrical power grid are difficult to avoid. The effect of AC electromagnetic field can be estimated and canceled [5] but intermittent disturbances caused by robot motors and electric currents degrade the velocity estimation accuracy. This way of localization is thus more suitable for human rescuers as a hand-held or wearable technology at this point.

7.2. Localization in smoke and fire

Promising technology is being developed in the project *SmokeBot*[1] whose goal is deployment of search and rescue robots in areas where fire and smoke are present. The sensors they use for localization are – apart from the common camera and laser alternative – a thermal imager and a compact ultra-wide-bandwidth radar. While angular resolution of the radar is still inferior to laser range-finders, it is perfectly appropriate for situations where smoke or dust prevents the laser-based sensor from functioning properly (it either detects obstacle at the point of transition between clean air and smoke/dust or measures nothing due to laser beam dissipation in the aerosol). Radar, on the other hand, penetrates aerosols and detects solid obstacles because of its much higher wavelength. This work is presented in [11, 12].

The thermal imager can be used to localize hot-spots, which are potential source of fire, but also to see through smoke. It can also be used to see through smoke, unlike

[1] European Union H2020 project SmokeBot (No. 645101; *http://www.smokebot.eu*)

visible light cameras. This feature could be used for localization as well, as works of [17, 3] suggest.

7.3. Multi-sensor fusion for localization

Data fusion for laser range-finder, wheel odometry, IMU and wireless signal strength has been shown in [2]. In their approach, 2D metric map of the environment is known beforehand. Thus the absolute position is observable for the localization system. The system is similarly to ours based on the EKF and implemented on a wheeled skid-steering robot. The signal-strength modality is used to limit computational complexity of the system by providing a rough initial guess of the robot position. In our work, we have focused on full 6-DOF localization which is required for search and rescue environments. We therefore had to consider roll and pitch angles and one extra position coordinate (elevation).

Another field with vast number of possible approaches and algorithms is visual-inertial localization and mapping. This modality often involves data fusion in form of tight coupling of IMU and camera. Examples and references to others can be found in [13, 18, 9]. Generally, visual-inertial localization is a mature technology being deployed in commercial application (e.g. augmented reality applications for cell phones heavily rely on it). Together with closely related field of localization and mapping based on RGB-D cameras[2], these techniques offer accurate localization requiring only cheap sensors. For search and rescue applications however, localization system has to offer back-up solution for situations when robot enters low-visibility areas, as we stress in our publications.

7.4. Metric point cloud maps

Point clouds and 3D occupancy grids are low-level representations of robot surrounding, useful for localization and mapping [21]. Technique used to match and align two similar point clouds is called Iterative Closest Point and has been the method of choice in our work in NIFTi and TRADR. Point clouds can be created directly by scanning robot surroundings by laser-range finders or reconstructed from series of digital images of the scanned object of environment. Work [14] shows how to merge point cloud map created by ground robot with a map reconstructed from aerial drone video.

Point cloud maps are suitable for traversability analysis and path planning. Results from the TRADR project [8] are one example. The higher-level information can be extracted from captured point clouds as well. Point cloud segments are classified and used to recognize previously visited places and to close loops in traversed maps in [7]. These results apply to search and rescue robotics as well as to the field of autonomous cars, which are often equipped with laser range-finders – results of [7] have been demonstrated on the *KITTI* dataset, which is intended for testing localization algorithms for autonomous cars.

7.5. Detecting changes in environment

In robotic research, environment is often considered static to simplify the problems being studied. However, dynamic changes in environment are inevitable in real world

[2]RGB-D camera is a classical digital camera combined with a depth sensor

and need to be addressed eventually [1, 10, 15, 23]. In search and rescue missions, changes of environment often indicate some sort of problem: collapsing structures, new obstacles the robot has to consider when planning and executing its motion. In different contexts however, changes in environment are natural process – people walk around, furniture is being moved, cars park at different spots. Modeling these processes can help when planning tasks the robot has to execute, e.g. where to go to find a given person at a given time [16].

7.6. Semantic mapping

Work [7] from the TRADR project consortium partially relates to this field by labeling point cloud segments and extracting set of features that allow later recognition. Advances in convolutional neural networks allow the fast segmentation and classification of images or point clouds. Giving low-level data human label makes human-robot interaction easier and allows high-level inference and planning. However in search and rescue field, the main problem with semantics we have faced was the very nature of the environment. Many classes of objects commonly present in contemporary datasets used for classifier training do not appear in conditions of earthquake aftermath or industrial accidents. If they do, they are often damaged and malformed. In our opinion, search and rescue robotics require their own ontologies and associated training datasets.

8. Conclusion

We have addressed the problem of robot localization for search and rescue missions in harsh sensory-deprived environments. Starting from basic proprioceptive sensor fusion for dead-reckoning localization, we have improved the system by integrating exteroceptive sensory modalities. The main contribution in the data fusion system has been comparing two state-of-the-art measurement models with our proposed one, which allows integration of low-sampling-rate sensory modalities. For tracked robots, we have proposed improved odometry which is suitable to harsh terrains where traversing vertical obstacles can be expected. These new models reduce localization drift and provide better motion estimates for higher-level mapping and planning systems.

We have explicitly addressed sensor failures and fail cases which can be expected in search and rescue conditions. Our findings have been experimentally verified with emphasis on real-world conditions and terrains. We have published our proposed algorithms and datasets in three publications and submitted one more; it is currently in review process.

In future work, our goal is to exploit tactile sensors for navigation in low-visibility conditions. Preliminary results suggest that it is possible to equip a tracked robot with force sensing through its rubber tracks. Identifying contact points between tracks and terrain can further improve motion estimation. Other promising modality are measurements of radio signal strength. Creating signal coverage model can be beneficial for path planning and autonomous return to safe areas when radio link is lost.

Bibliography

[1] H. Andreasson, M. Magnusson, and A. Lilienthal. Has somethong changed here? autonomous difference detection for security patrol robots. In *2007 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 3429–3435, Oct 2007. 70

[2] M. M. Atia, S. Liu, H. Nematallah, T. B. Karamat, and A. Noureldin. Integrated indoor navigation system for ground vehicles with automatic 3-d alignment and position initialization. *IEEE Transactions on Vehicular Technology*, 64(4):1279–1292, April 2015. 69

[3] Furkan Çakmak, Erkan Uslu, M Fatih Amasyalı, and Sırma Yavuz. Thermal based exploration for search and rescue robots. In *INnovations in Intelligent SysTems and Applications (INISTA), 2017 IEEE International Conference on*, pages 113–118. IEEE, 2017. 69

[4] David Caruso, Alexandre Eudes, Martial Sanfourche, David Vissière, and Guy le Besnerais. Robust indoor/outdoor navigation through magneto-visual-inertial optimization-based estimation. In *Intelligent Robots and Systems (IROS), 2017 IEEE/RSJ International Conference on*, pages 4402–4409. IEEE, 2017. 68

[5] C. I. Chesneau, M. Hillion, and C. Prieur. Motion estimation of a rigid body with an ekf using magneto-inertial measurements. In *2016 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, pages 1–6, Oct 2016. 68

[6] E. Dorveaux, T. Boudot, M. Hillion, and N. Petit. Combining inertial measurements and distributed magnetometry for motion estimation. In *Proceedings of the 2011 American Control Conference*, pages 4249–4256, June 2011. 68

[7] Renaud Dubé, Andrei Cramariuc, Daniel Dugas, Juan Nieto, Roland Siegwart, and Cesar Cadena. SegMap: 3d segment mapping using data-driven descriptors. In *Robotics: Science and Systems (RSS)*, 2018. 69, 70

[8] Renaud Dubé, Abel Gawel, Cesar Cadena, Roland Siegwart, Luigi Freda, and Mario Gianni. 3d localization, mapping and path planning for search and rescue operations. In *Safety, Security, and Rescue Robotics (SSRR), 2016 IEEE International Symposium on*, pages 272–273. IEEE, 2016. 69

[9] Jakob Engel, Thomas Schöps, and Daniel Cremers. Lsd-slam: Large-scale direct monocular slam. In *European Conference on Computer Vision*, pages 834–849. Springer, 2014. 69

[10] F. Ferri, M. Gianni, M. Menna, and F. Pirri. Dynamic obstacles detection and 3d map updating. In *2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 5694–5699, Sept 2015. 70

[11] P. Fritsche, B. Zeise, P. Hemme, and B. Wagner. Fusion of radar, lidar and thermal information for hazard detection in low visibility environments. In *2017 IEEE*

International Symposium on Safety, Security and Rescue Robotics (SSRR), pages 96–101, Oct 2017. 68

[12] Paul Fritsche, Simon Kueppers, Gunnar Briese, and Bernardo Wagner. Radar and lidar sensorfusion in low visibility environments. In *ICINCO (2)*, pages 30–36, 2016. 68

[13] Jorge Fuentes-Pacheco, José Ruiz-Ascencio, and Juan Manuel Rendón-Mancha. Visual simultaneous localization and mapping: a survey. *Artificial Intelligence Review*, 43(1):55–81, 2015. 69

[14] A. Gawel, R. Dub, H. Surmann, J. Nieto, R. Siegwart, and C. Cadena. 3d registration of aerial and ground robots for disaster response: An evaluation of features, descriptors, and transformation estimation. In *2017 IEEE International Symposium on Safety, Security and Rescue Robotics (SSRR)*, pages 27–34, Oct 2017. 69

[15] Daniel Girardeau-Montaut, Michel Roux, Raphaël Marc, and Guillaume Thibault. Change detection on points cloud data acquired with a ground laser scanner. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 36(part 3):W19, 2005. 70

[16] Tomáš Krajník, Jaime P Fentanes, Joao M Santos, and Tom Duckett. Fremen: Frequency map enhancement for long-term mobile robot autonomy in changing environments. *IEEE Transactions on Robotics*, 33(4):964–977, 2017. 70

[17] Guoyu Lu, Yan Yan, Li Ren, Philip Saponaro, Nicu Sebe, and Chandra Kambhamettu. Where am i in the dark: Exploring active transfer learning on the use of indoor localization based on thermal imaging. *Neurocomputing*, 173:83–92, 2016. 69

[18] R. Mur-Artal, J. M. M. Montiel, and J. D. Tards. Orb-slam: A versatile and accurate monocular slam system. *IEEE Transactions on Robotics*, 31(5):1147–1163, Oct 2015. 69

[19] Robin R. Murphy. *Disaster Robotics*. The MIT Press, 2014. 2, 3

[20] Keiji Nagatani, Seiga Kiribayashi, Yoshito Okada, Kazuki Otake, Kazuya Yoshida, Satoshi Tadokoro, Takeshi Nishimura, Tomoaki Yoshida, Eiji Koyanagi, Mineo Fukushima, et al. Emergency response to the nuclear accident at the fukushima daiichi nuclear power plants using mobile rescue robots. *Journal of Field Robotics*, 30(1):44–63, 2013. 2

[21] Sajad Saeedi, Michael Trentini, Mae Seto, and Howard Li. Multiple-robot simultaneous localization and mapping: A review. *Journal of Field Robotics*, 33(1):3–46, 2016. 69

[22] Sebastian Thrun, Wolfram Burgard, and Dieter Fox. *Probabilistic Robotics*. The MIT Press, 2005. 2

[23] L. Wellhausen, R. Dub, A. Gawel, R. Siegwart, and C. Cadena. Reliable real-time change detection and mapping for 3d lidars. In *2017 IEEE International Symposium on Safety, Security and Rescue Robotics (SSRR)*, pages 81–87, Oct 2017. 70

A. Publications of the author

A.1. Thesis-related articles in peer-reviewed journals with impact factor

[A.1] J. Šimánek, M. Reinstein and V. Kubelka. Evaluation of the EKF-Based Estimation Architectures for Data Fusion in Mobile Robots. In *IEEE/ASME Transactions on Mechatronics*, 2015. **18 citations in WoS**.

[A.2] V. Kubelka et al. Robust Data Fusion of Multimodal Sensory Information for Mobile Robots. In *Journal of Field Robotics*, 2015. **8 citations in WoS**.

[A.3] J. Šimánek, V. Kubelka and M. Reinstein. Improving multi-modal data fusion by anomaly detection. In *Autonomous Robots*, 2015. **5 citations in WoS**.

[A.4] V. Kubelka, M. Reinstein and T. Svoboda. Improving multimodal data fusion for mobile robots by trajectory smoothing. In *Robotics and Autonomous Systems*, 2016. **1 citation in WoS**

A.2. Thesis-related conference proceedings indexed in WoS

[A.5] V. Kubelka and M. Reinstein. Complementary filtering approach to orientation estimation using inertial sensors only. In *2012 IEEE International Conference on Robotics and Automation*, 2012. **18 citations in WoS**.

[A.6] M. Reinstein, V. Kubelka and K. Zimmermann. Terrain adaptive odometry for mobile skid-steer robots. In *2013 IEEE International Conference on Robotics and Automation*, 2013. **8 citations in WoS**.

[A.7] V. Kubelka and M. Reinstein. Combining Complementary Motion Estimation Approaches to Increase Reliability in Urban Search & Rescue Missions. In *International Workshop on Modelling and Simulation for Autonomous Systems*, 2014.

[A.8] I. Kruijff-Korbayová et al. Deployment of ground and aerial robots in earthquake-struck Amatrice in Italy (brief report). In *2016 IEEE International Symposium on Safety, Security, and Rescue Robotics (SSRR)*, 2016.

A.3. Other conference proceedings

[A.9] V. Šalanský, V. Kubelka, K. Zimmermann, M. Reinstein, and T. Svoboda. Touching without vision: terrain perception in sensory deprived environments. In *Proceedings of the 21st Computer Vision Winter Workshop*, 2016. Authorship of V. Kubelka 10%.

[A.10] M. Jirků, V. Kubelka, and M. Reinstein. WiFi localization in 3D. In 2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2016. **1 citation in WoS**