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. © 2014 Wiley Periodicals, Inc.

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

View this article online at wileyonlinelibrary.com • DOI: 10.1002/rob.21535

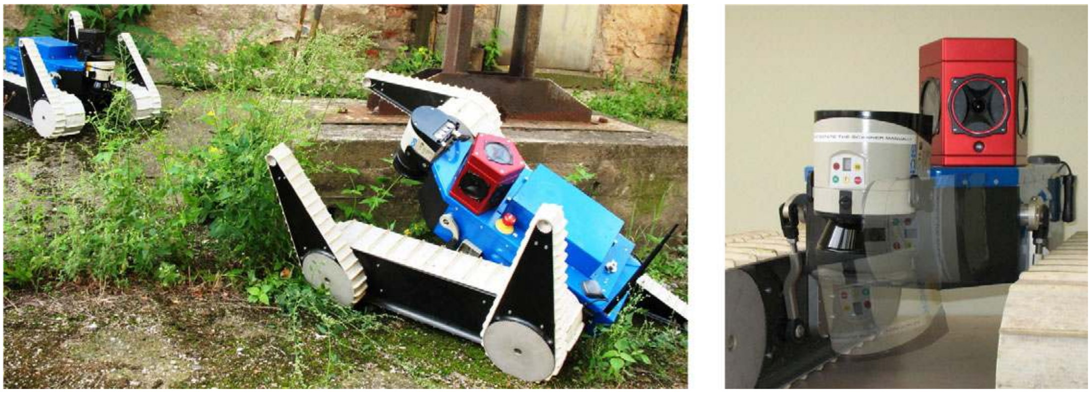

Figure 1. Picture of two USAR UGVs used for experimental evaluation (FP7-ICT-247870 NIFTi project) and a detail of the sensor setup (a PointGrey Ladybug 3 omnicamera and a rotating SICK LMS-151 laser rangefinder). See Section 3.1 for more details.

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]

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

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

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°
	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

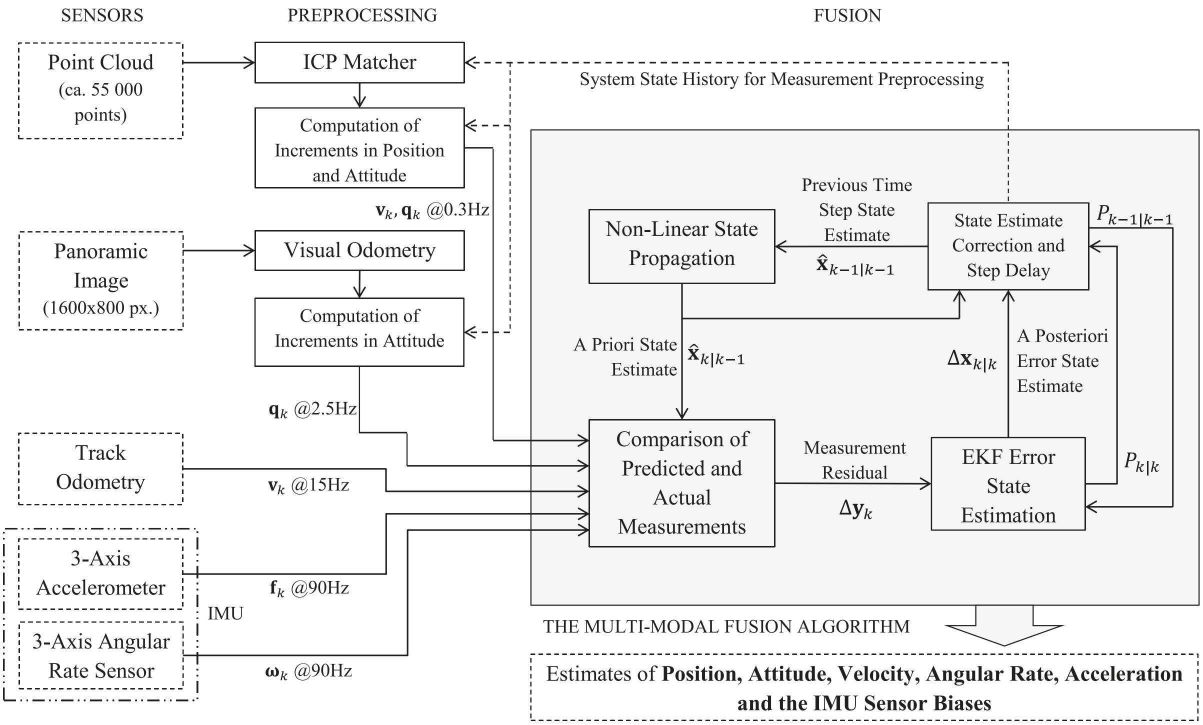
SENSORS
PREPROCESSING
FUSION
Point Cloud (ca. 55 000 points)
ICP Matcher
System State History for Measurement Preprocessing
Computation of Increments in Position and Attitude
$\mathbf{v}_k, \mathbf{q}_k$ @0.3Hz
Panoramic Image (1600x800 px.)
Visual Odometry
Computation of Increments in Attitude
$\mathbf{q}_k$ @2.5Hz
Track Odometry
$\mathbf{v}_k$ @15Hz
3-Axis Accelerometer
IMU
$\mathbf{f}_k$ @90Hz
3-Axis Angular Rate Sensor
$\boldsymbol{\omega}_k$ @90Hz
Non-Linear State Propagation
Previous Time Step State Estimate
$\hat{\mathbf{x}}_{k-1|k-1}$
State Estimate Correction and Step Delay
$P_{k-1|k-1}$
A Priori State Estimate
$\hat{\mathbf{x}}_{k|k-1}$
$\Delta\mathbf{x}_{k|k}$
A Posteriori Error State Estimate
Comparison of Predicted and Actual Measurements
Measurement Residual
$\Delta\mathbf{y}_k$
EKF Error State Estimation
$P_{k|k}$
THE MULTI-MODAL FUSION ALGORITHM
Estimates of **Position, Attitude, Velocity, Angular Rate, Acceleration and the IMU Sensor Biases**

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:

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

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

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

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

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

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

The noise-coupling matrix describes how particular noise terms affect the system state. Each $\mathbf{n}_{(\cdot)}$ 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, $Q_c = \operatorname{diag}(\sigma^2_{\theta_x}, \sigma^2_{\theta_y}, \sigma^2_{\theta_z}, \sigma^2_{v_x}, \sigma^2_{v_y}, \ldots)$, 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,

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

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 $\Delta\mathbf{x}$.

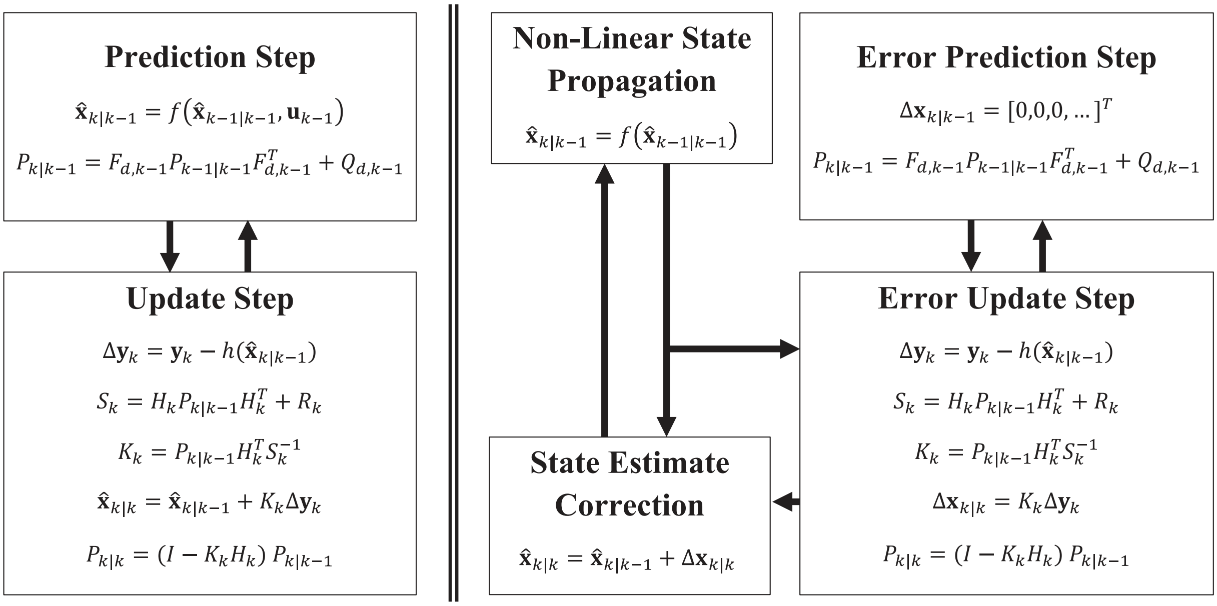
Prediction Step
$\hat{\mathbf{x}}_{k|k-1} = f(\hat{\mathbf{x}}_{k-1|k-1}, \mathbf{u}_{k-1})$
$P_{k|k-1} = F_{d,k-1}P_{k-1|k-1}F_{d,k-1}^T + Q_{d,k-1}$
Update Step
$\Delta\mathbf{y}_k = \mathbf{y}_k - h(\hat{\mathbf{x}}_{k|k-1})$
$S_k = H_k P_{k|k-1} H_k^T + R_k$
$K_k = P_{k|k-1} H_k^T S_k^{-1}$
$\hat{\mathbf{x}}_{k|k} = \hat{\mathbf{x}}_{k|k-1} + K_k \Delta\mathbf{y}_k$
$P_{k|k} = (I - K_k H_k)\, P_{k|k-1}$
Non-Linear State Propagation
$\hat{\mathbf{x}}_{k|k-1} = f(\hat{\mathbf{x}}_{k-1|k-1})$
State Estimate Correction
$\hat{\mathbf{x}}_{k|k} = \hat{\mathbf{x}}_{k|k-1} + \Delta\mathbf{x}_{k|k}$
Error Prediction Step
$\Delta\mathbf{x}_{k|k-1} = [0,0,0,\dots]^T$
$P_{k|k-1} = F_{d,k-1}P_{k-1|k-1}F_{d,k-1}^T + Q_{d,k-1}$
Error Update Step
$\Delta\mathbf{y}_k = \mathbf{y}_k - h(\hat{\mathbf{x}}_{k|k-1})$
$S_k = H_k P_{k|k-1} H_k^T + R_k$
$K_k = P_{k|k-1} H_k^T S_k^{-1}$
$\Delta\mathbf{x}_{k|k} = K_k \Delta\mathbf{y}_k$
$P_{k|k} = (I - K_k H_k)\, P_{k|k-1}$

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^I_R 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

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

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':

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

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

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

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:

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

which can be expressed in the matrix form

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

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^O_R:

$$\mathbf{v}_O = C^O_R \mathbf{v}_R, \tag{25}$$

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^O_R as constant—its value was obtained by means of calibration. The measurement equation is then as follows:

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

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

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

and expressed in the matrix form

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

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:

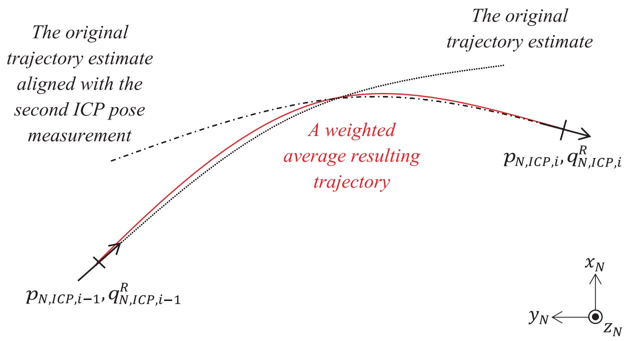
The original trajectory estimate aligned with the second ICP pose measurement
The original trajectory estimate
A weighted average resulting trajectory
$p_{N,ICP,i}, q^R_{N,ICP,i}$
$p_{N,ICP,i-1}, q^R_{N,ICP,i-1}$
x_N
y_N
z_N

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 $\delta\boldsymbol{\theta}_{\mathrm{ICP},i}$,

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

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 $\delta\boldsymbol{\theta}_{\mathrm{VO},i}$,

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

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.

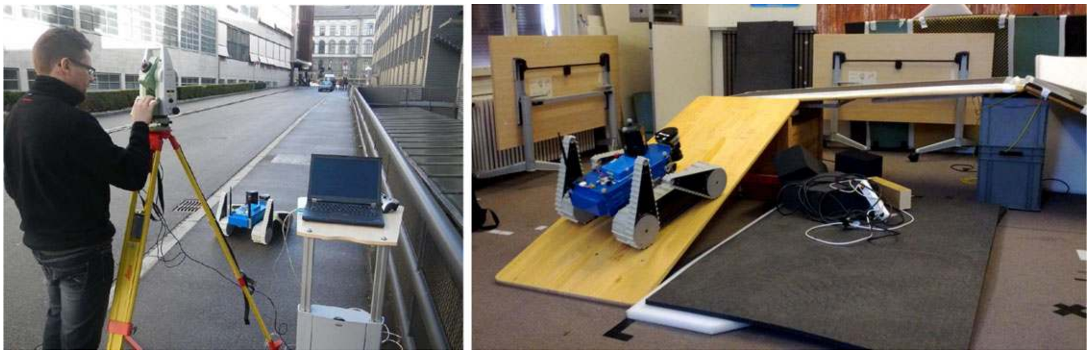

Figure 5. The experimental setup with the Leica reference theodolite for obtaining ground truth trajectory (left). Part of the 3D semistructured environment for an indoor test with motion capture ground truth (right).

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.

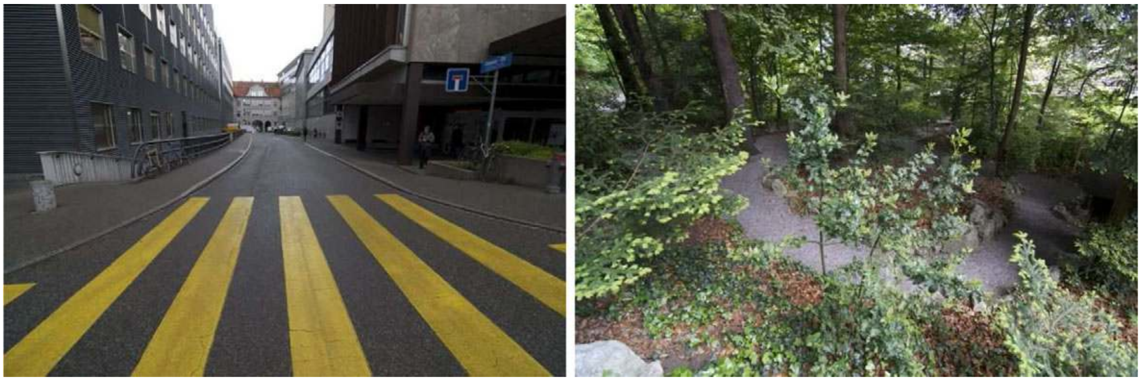

Figure 6. Pictures of the outdoor environments in Zurich. Left: street canyon, right: urban park.

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.

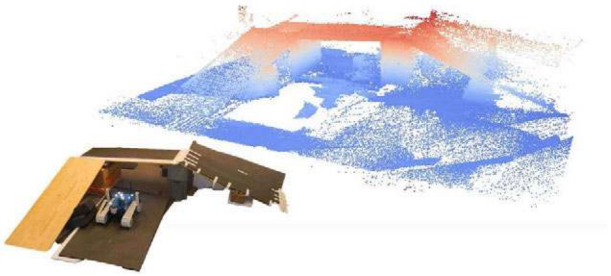

Figure 7. The 3D structure for testing obstacle traversability shown as a metric map created by ICP.

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

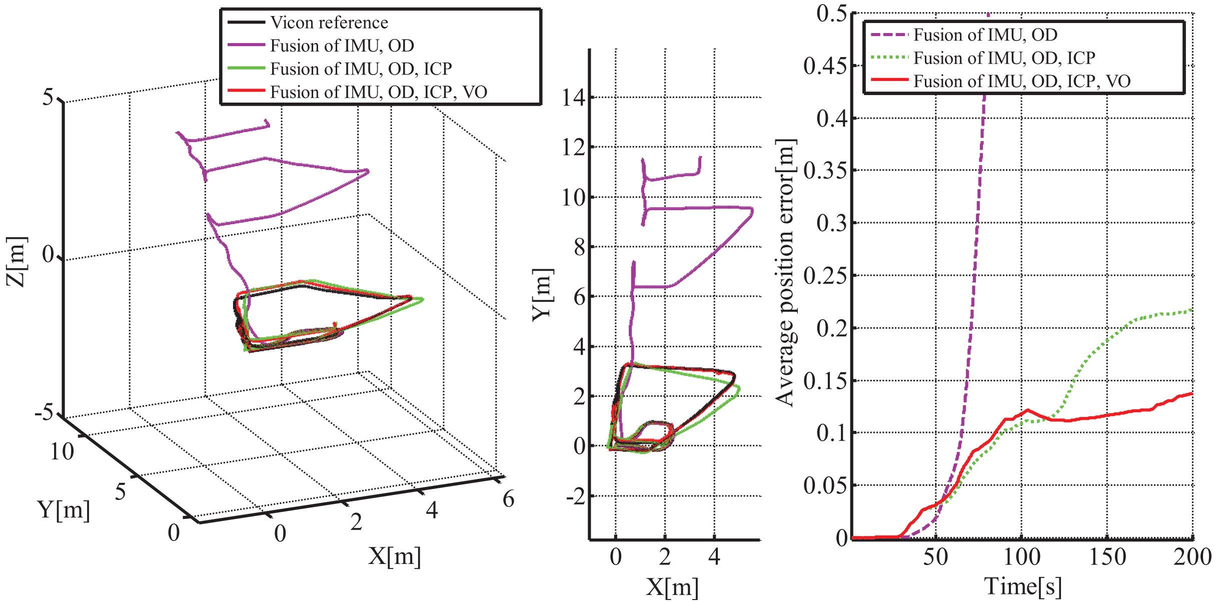

Figure 8. Trajectories obtained by fusing different combinations of modalities during the indoor experiment testing obstacle (depicted in Figure 7) traversability under high slippage (left, middle); development of the average position error (right).

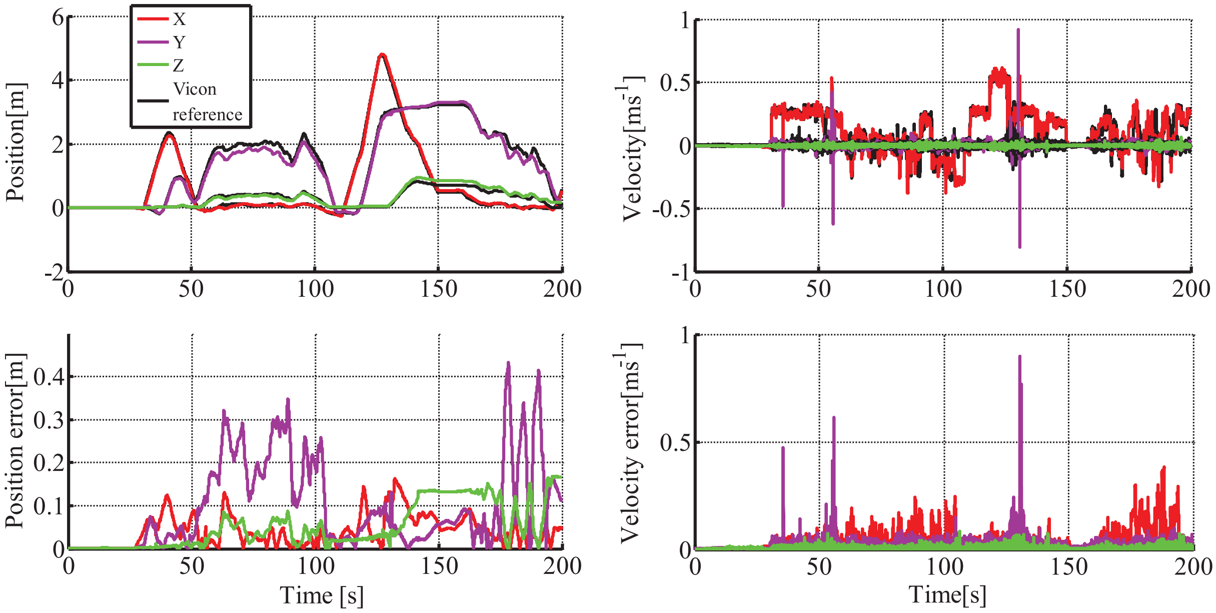

Figure 9. The corrected position (top left) and velocity estimates (top right) for the IMU+OD+ICP+VO combination corresponding to the trajectory in Figure 8 (testing obstacle traversability). Errors in position and velocity are obtained as the norm of difference between the Vicon reference and the corresponding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

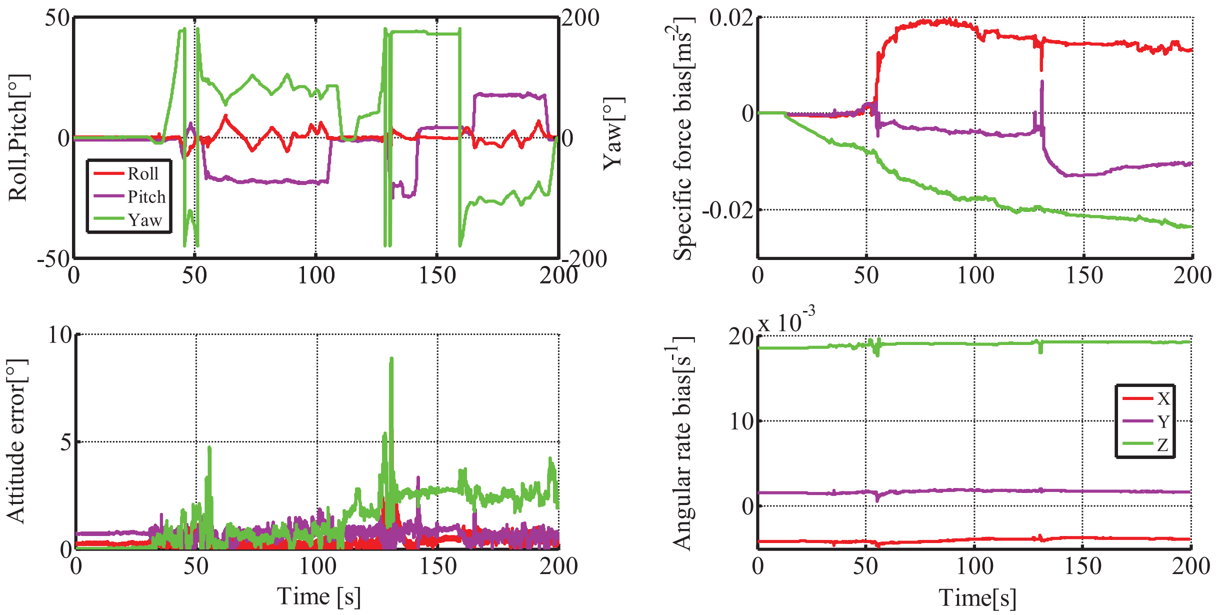

Figure 10. The corrected attitude estimates (top left) for the full multimodal combination IMU+OD+ICP+VO corresponding to the trajectory shown in Figure 8 (testing obstacle traversability). Errors in attitude are obtained as the difference between the Vicon reference and the corresponding state at each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

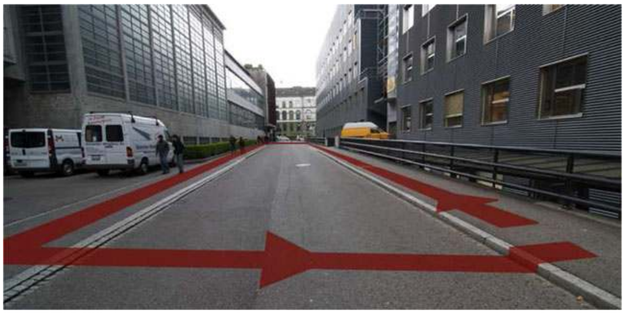

Figure 11. An example of trajectory driven by the robot over the Clausiusstrasse street.

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

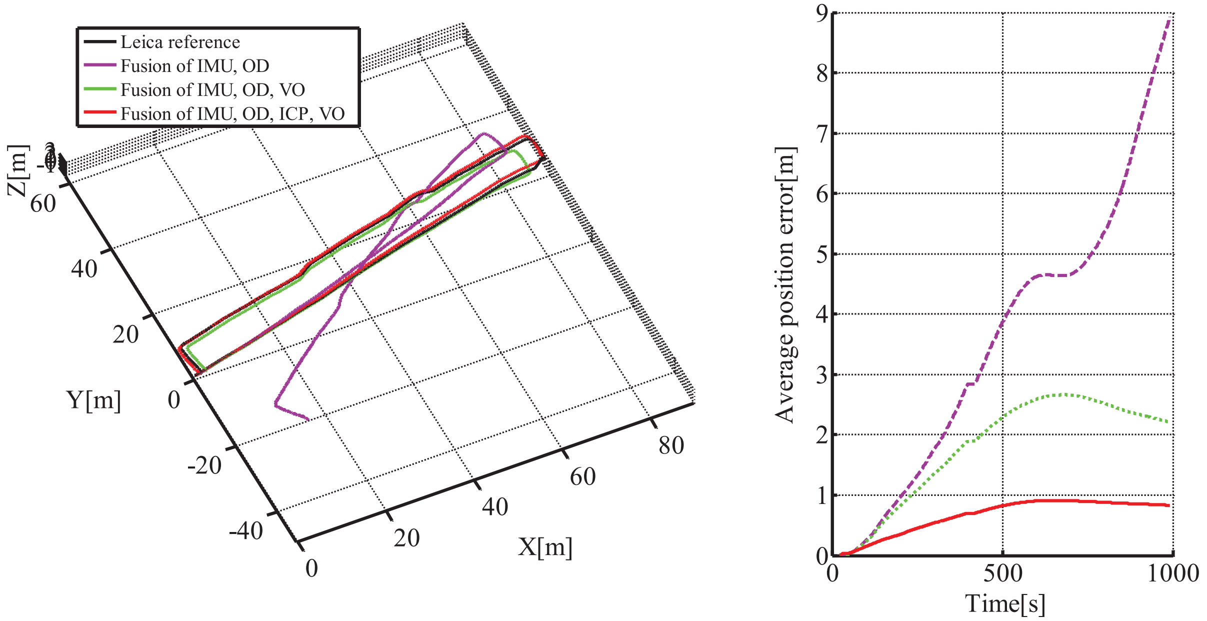

Figure 12. Trajectories obtained by fusing different combinations of modalities during the outdoor experiment with Leica reference system (left) and the corresponding average position error in time (right).

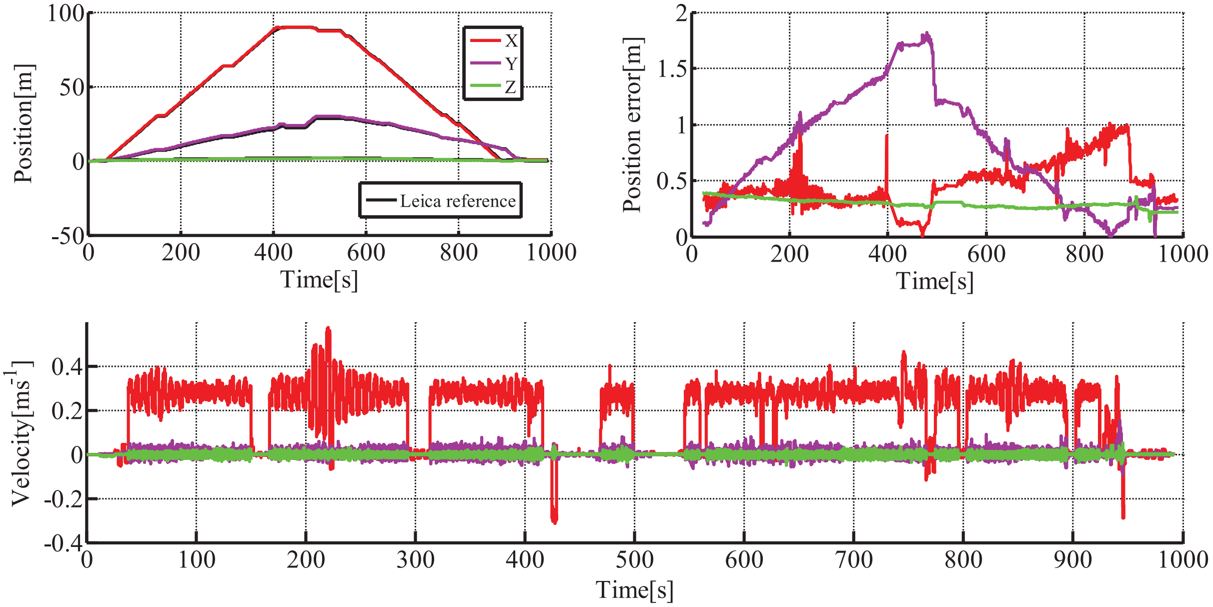

Figure 13. The position and velocity estimates (top left and bottom, respectively) for the IMU+OD+ICP+VO combination corresponding to the outdoor trajectory in Figure 12; errors in position obtained as the norm of differences between the Leica reference and the corresponding state at each time-step (top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

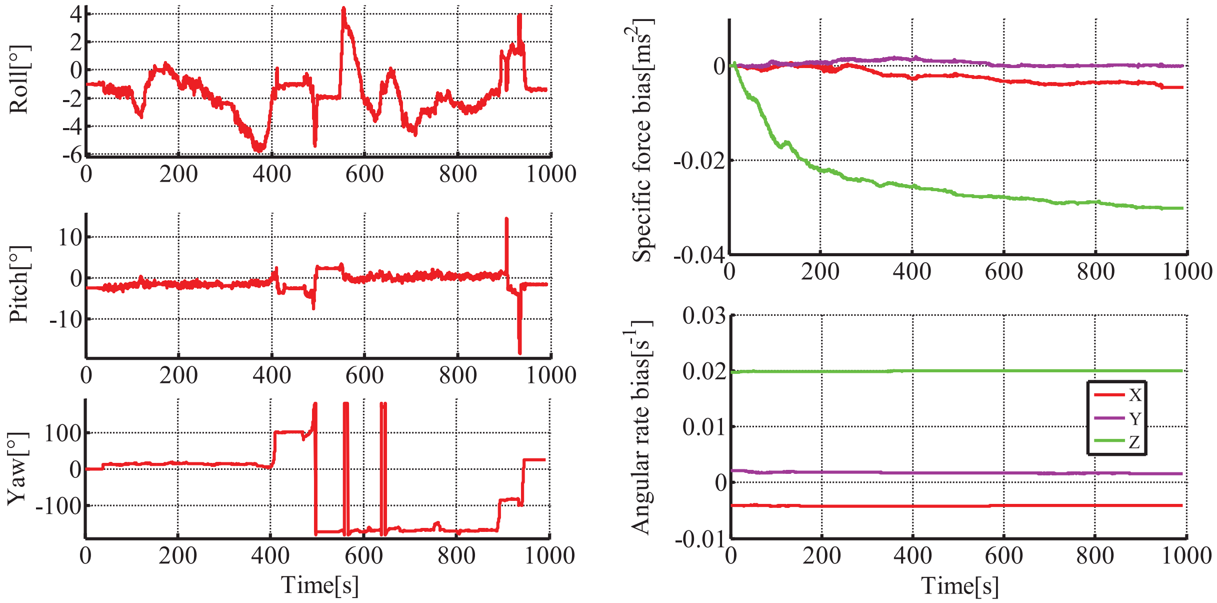

Figure 14. The attitude estimates (left) for the IMU+OD+ICP+VO combination corresponding to the outdoor trajectory in Figure 12; biases estimated for the specific forces (top right) and angular rates (bottom right).

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.

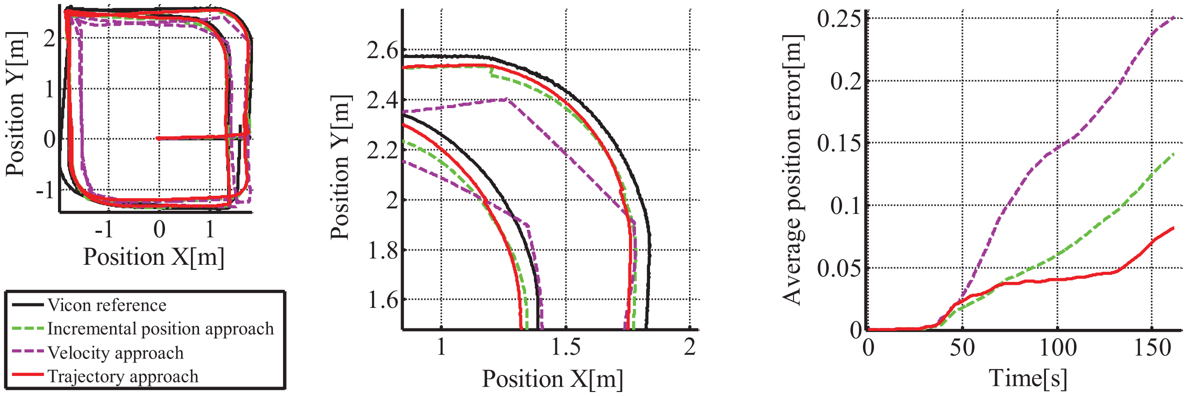

Figure 15. Comparison of effects of the three different ICP aiding approaches on the estimated trajectory (left, middle) and on the average position error (right). Note the *corner cutting* effect of the *velocity approach*.

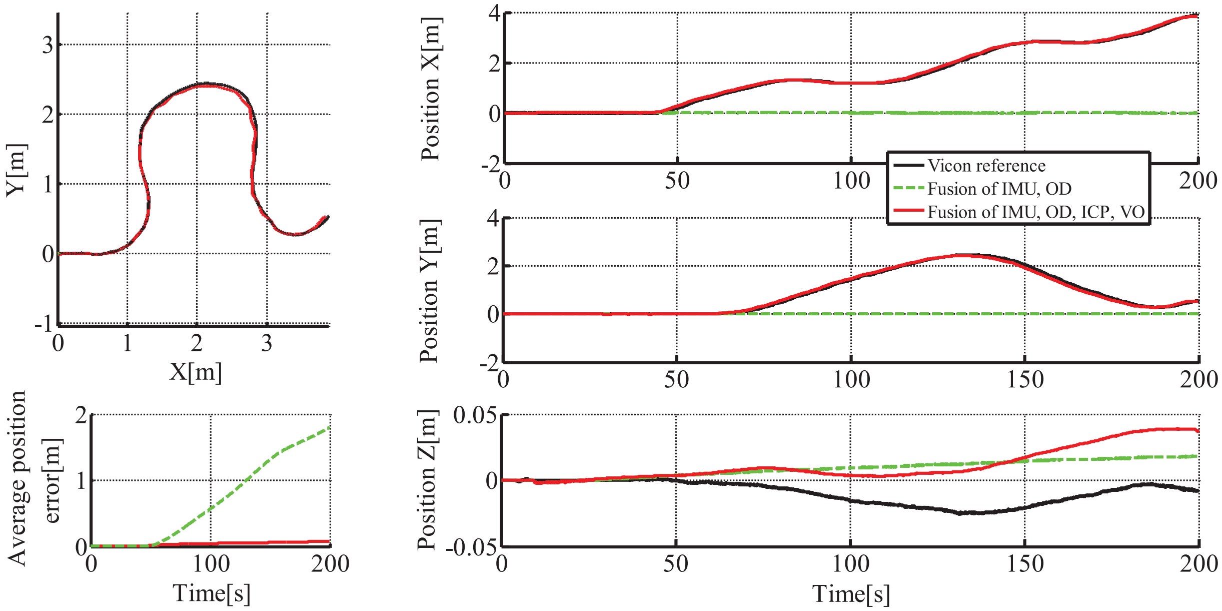

Figure 16. Test trajectory for robot slippage. Black line: ground truth; red solid line: state estimate with all four modalities; green dashed line: IMU and odometry fusion. Top left: top view of the trajectory; bottom left: average error as a function of time; top, middle, bottom right: evolution of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

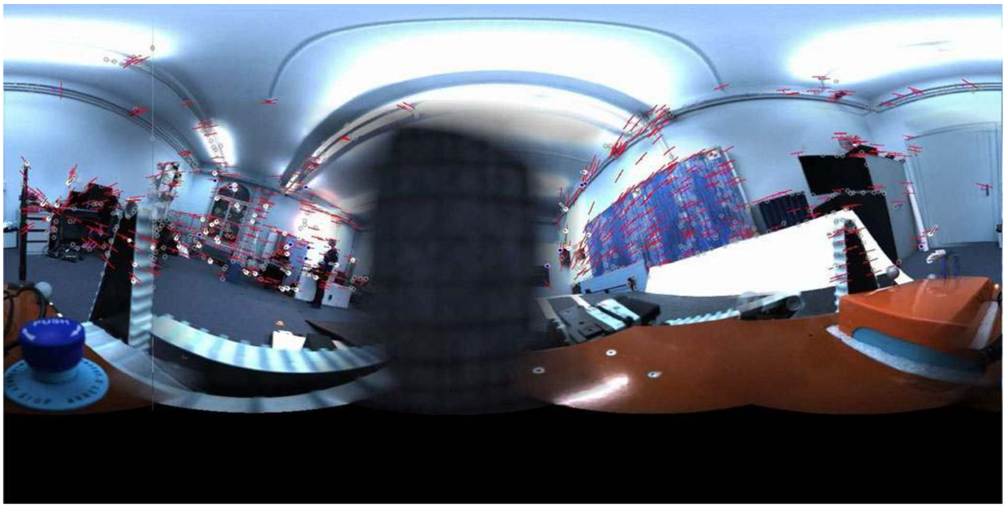

Figure 17. Picture from the partially occluded omnicamera. Notice the dark rectangle in the middle.

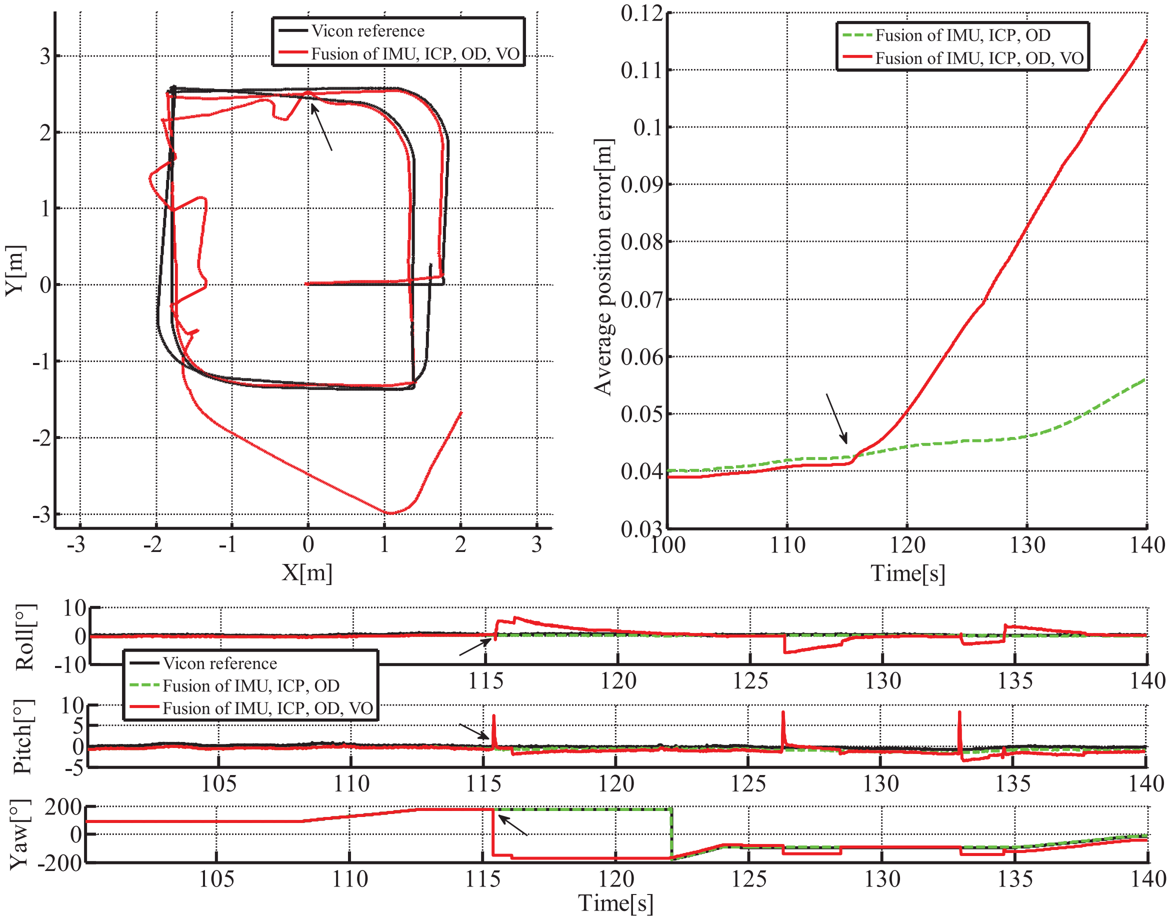

Figure 18. Trajectory reconstruction with several faulty VO motion estimates. Black line: ground truth; solid red line: state estimate with all four modalities; dashed green line: state estimate excluding visual odometry; black arrow: visual odometry failure. Top left: top view of the trajectory; top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

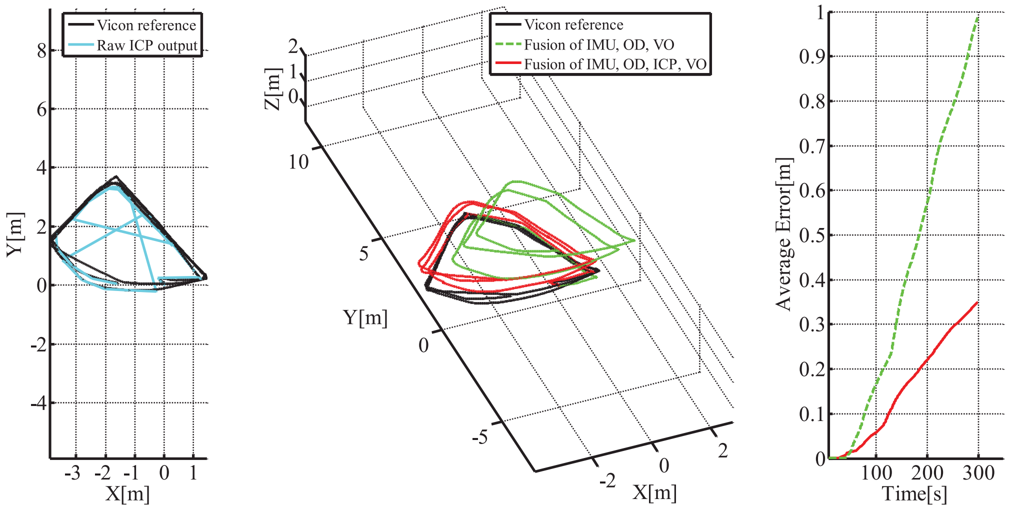

Figure 19. Trajectory estimates in the case of low ICP frequency. Black line: ground truth; cyan line: positions estimated by ICP alone; red line: state estimate with all four modalities; green dashed line: state estimate excluding ICP measurements. Left: top view; middle: 3D view; right: average position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

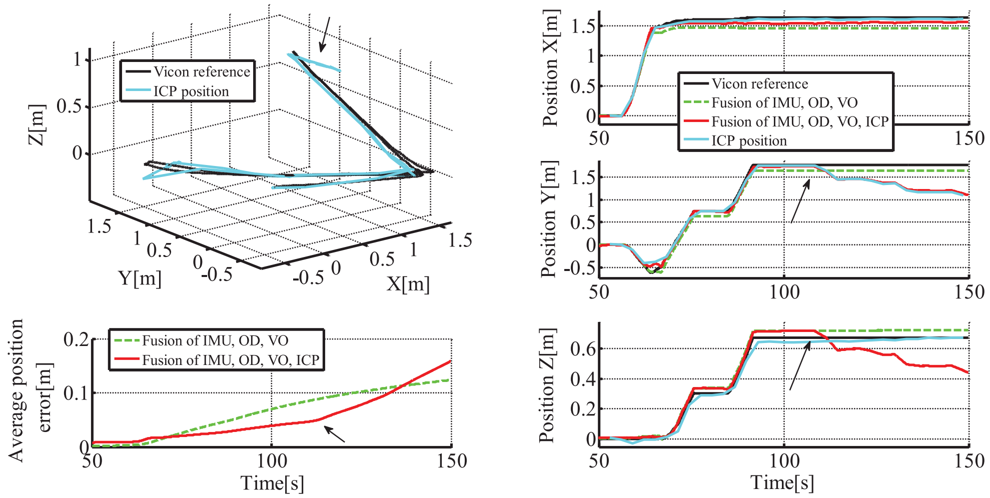

Figure 20. Trajectory estimates in the case of a moving obstacle in a reduced field of view. Solid black line: ground truth; solid red line: state estimate with all four modalities; dashed green line: state estimate excluding ICP measurements; cyan line: position estimated by ICP alone; black arrow: 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

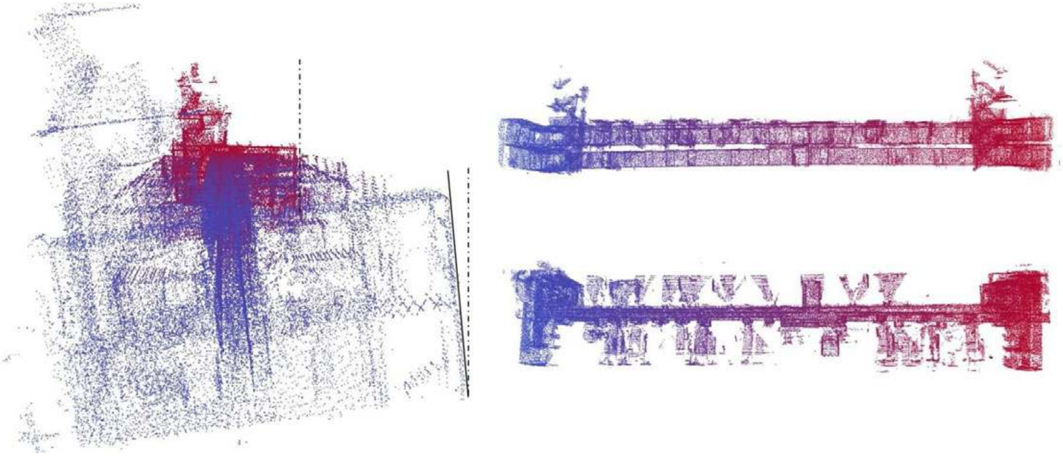

Figure 21. Deformed point cloud map created by ICP. The points are colored alongside the corridor from red (initial position) to blue. Left: front view; top right: side view; bottom right: top view.

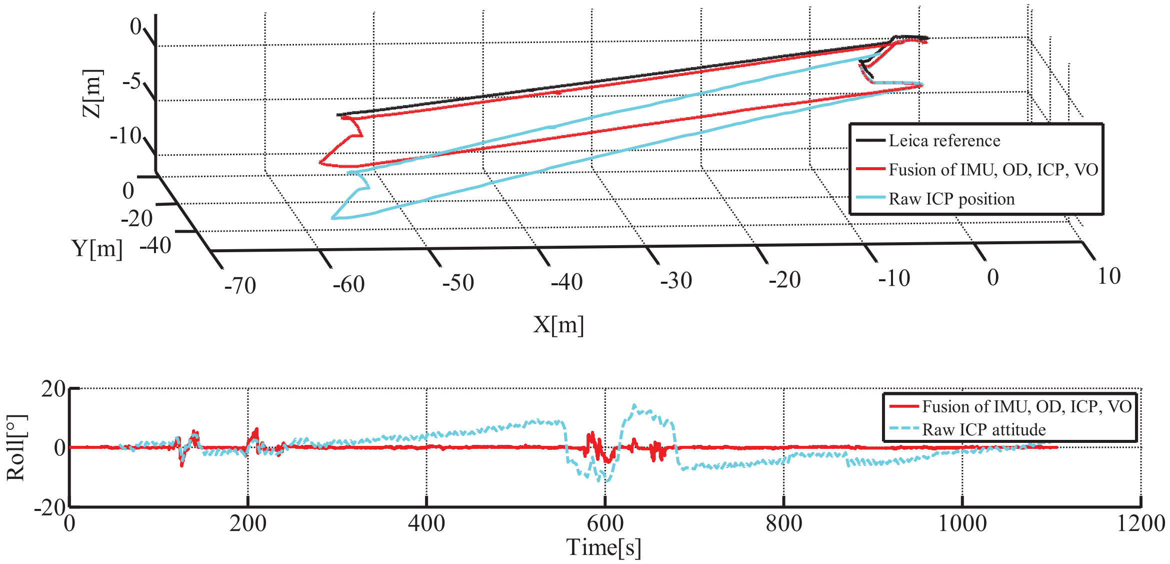

Figure 22. Trajectory estimates in the case of map deformation. Solid black line: ground truth; solid red line: state estimate with all four modalities; cyan line: position estimated by ICP alone. Top: side view; bottom: roll angle along the trajectory.

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).