Czech Technical University in Prague
Faculty of Electrical Engineering

Doctoral Thesis

October 2015

Jakub Šimánek

Czech Technical University in Prague
Faculty of Electrical Engineering
Department of Measurement

Data fusion for localization using state estimation and machine learning

Doctoral Thesis

Jakub Šimánek

Prague, October 2015

Ph.D. Programme: Electrical Engineering and Information Technology
Branch of study: Air Traffic Control

Supervisor: doc. Ing. Jan Roháč, Ph.D.
Supervisor-Specialist: Ing. Michal Reinštein, Ph.D.

Thesis Supervisor:

doc. Ing. Jan Roháč, Ph.D.
Department of Measurement
Faculty of Electrical Engineering
Czech Technical University in Prague
Technická 1902/2
160 00 Prague 6
Czech Republic

Thesis Supervisor-Specialist:

Ing. Michal Reinštein, Ph.D.
Department of Cybernetics
Faculty of Electrical Engineering
Czech Technical University in Prague
Karlovo náměstí 13
121 35 Prague 2
Czech Republic

Abstract

Using multiple sensory information is acknowledged as one of the major topics in the navigation of aerial and ground vehicles. This doctoral thesis considers localization as a state estimation problem, which is solved by data fusion techniques and supported by machine learning methods. It attempts to address the issue of developing a better localization system for a ground vehicle by seeking the best possible pose estimator (i.e., position, velocity and attitude) and improving its robustness to unexpected sensor measurements. The vehicle of interest is represented by a skid-steer tracked mobile robot; however, all the algorithms work with a sensory set, which can be with minor changes deployed on any vehicle, legged, wheeled, tracked, or aerial. First part of this thesis explores the development of different state estimation architectures, which exploit the extended Kalman filter for full 3D dead reckoning (i.e., incremental or relative pose estimation). The purpose of this part is to use inertial and odometry dead reckoning to its optimal extent, considering both the performance and computational complexity. Such combination of proprioceptive sensory modalities used on a ground vehicle is expected to provide the core localization—foundation for any other higher level localization or navigation systems. Second part of the dissertation investigates means of improving overall robustness and performance of the multi-modal state estimation. Different sensory modalities are prone to various types of errors, especially in an environment that changes dynamically. Therefore, the thesis shows the importance of identifying and rejecting unexpected or erroneous measurements. The multi-modal data fusion is based on inertial and odometry measurements aided by information from a camera and laser range finder. These two exteroceptive modalities are in particular prone to real-world disturbances, therefore, they are the subject of anomaly detection process. Various state-of-the-art machine learning methods (i.e., logistic regression, Support Vector Machines, Gaussian Mixture Models and Gaussian Processes) are applied in a Kalman filter framework to monitor the measurements and overcome the commonly used covariance monitoring and chi-squared gating test. Verification of all the techniques in this thesis is supported by extensive experimental datasets collected with a real mobile robot in both indoor and challenging outdoor environments.

Keywords:

state estimation, data fusion, mobile robot, localization, Kalman filter, anomaly detection, machine learning.

Abstrakt

Využití dat z různých typů senzorů je považováno za jedno z hlavních témat řešených v problému navigace pozemních a vzdušných prostředků. Hlavním cílem dizertační práce je vývoj lokalizačního systému, který využívá fúze senzorických dat, stavového odhadu a metod strojového učení. Práce je zaměřena na vývoj algoritmů, které určují pozici, rychlost a orientaci v prostoru, a zvýšení jejich robustnosti vůči neočekávaným nebo chybným měřením. Ačkoliv je cílová platforma reprezentována pozemním pásovým robotem, algoritmy využívají množinu senzorů, která je aplikovatelná na různých typech prostředků, pozemních i vzdušných. První část práce je zaměřena na vývoj algoritmů pro odhad stavu využívajících rozšířeného Kalmanova filtru pro plně prostorovou inkrementální lokalizaci (tzv. dead-reckoning). Cílem je optimálně zkombinovat inerciální data a odometrii s přihlédnutím jak k přesnosti lokalizace, tak k výpočetní náročnosti. Od takové kombinace proprioceptivních senzorů je předpokládána základní lokalizace poskytující podklad pro vyšší úrovně navigačního systému. Druhá část práce se zabývá způsoby vylepšení robustnosti a přesnosti lokalizace. Různé zdroje senzorických dat jsou náchylné na různé typy chyb, obzvláště v dynamicky se měnícím prostředí. Práce se v druhé části zaměřuje na způsoby identifikace a následného vynechání chybných anomálních měření z kamery a laserového dálkoměru. Exteroceptivní data jsou často významně ovlivněna vnějším okolím, které způsobuje neočekávaná nebo chybná měření, a proto je nutné taková měření monitorovat. V práci je zkoumána kombinace Kalmanova filtru s různými metodami strojového učení, které jsou aplikovány jako klasifikátory anomálních dat. Tyto kombinace jsou porovnány s technikami běžně využívanými pro monitorování Kalmanova filtru. Navržený systém je důkladně testován a ověřen v reálných experimentech provedených v náročných prostředích, jako jsou městské zástavby, venkovní prostory a poničené oblasti napodobující průběh záchranných misí.

Klíčová slova:

stavová estimace, fúze dat, mobilní robotika, lokalizace, Kalmanův filtr, detekce anomálií, strojové učení.

Acknowledgements

I would like to thank my supervisors Jan Roháč and Michal Reinštein for their excellent guidance, necessary resources and support, patience and numerous discussions, that inspired me in many ways during the last four years.

Special thanks go to the staff of the Department of Measurement for the pleasant and flexible environment for my research, and to my colleagues from the Laboratory of Aircraft Instrumentation Systems for creating the perfect research environment and unforgettable friendships. I also have to thank Vladimír Kubelka, my colleague from the Department of Cybernetics, for his invaluable cooperation, insights and his help in carrying out hundreds of experiments and sharing his valuable datasets.

During my time as a PhD student, I had the chance to visit the Artificial Intelligence Laboratory, University of Zurich, lead by Prof. Rolf Pfeifer. I want to thank Matěj Hoffmann, who offered me the opportunity of the internship and who gave me the chance to cooperate with him in the field of embodied robotics. My stay at the laboratory was a memorable experience filled with everyday inspiration.

Finally and most importantly, I am deeply grateful to my family and my dearest Monika for their continuous support and patience.

This work would not have been possible without the financial support. Therefore, I acknowledge the Grant Agency of the CTU in Prague (SGS15/163/OHK3/2T/13, SGS13/144/OHK3/2T/13).

Contents

List of Figures

List of Tables

Abbreviations

2D / 3D	Two / Three Dimensions
DOF	Degree of Freedom
UGV	Unmanned Ground Vehicle
UAV	Unmanned Aerial Vehicle
KF / EKF / UKF	Kalman Filter / Extended / Unscented
CF	Complementary Filter
AD	Anomaly Detection
FDI	Fault Detection and Identification
GPS	Global Positioning System
IMU	Inertial Measurement Unit
VO	Visual Odometry
SLAM	Simultaneous Localization and Mapping
RTC	Robotic Terrain Classification
ICP	Iterative Closest Points
NN	Neural Network
PCA	Principal Component Analysis
FFT	Fast Fourier Transform
GMM	Gaussian Mixture Model
SVM	Support Vector Machine
REG	Logistic Regression
COV	Covariance Analysis
CHI	Chi-squared test
GP	Gaussian Process
SE	Squared Exponential (GP kernel)
RQ	Rational Quadratic (GP kernel)
FITC	Fully Independent Training Conditional (GP approximation)
LA	Laplace Approximation
EP	Expectation Propagation
MCMC	Markov Chain Monte Carlo (sampling)
TP / TPR	True Positive / True Positive Rate
FP	False Positive
TN / TNR	True Negative / True Negative Rate
FN	False Negative
PR	Precision
ROC	Receiver Operating Characteristic
AUC	Area Under Curve

Chapter 1

Introduction

In ground and aerial mobile robotics, localization is defined as the *problem of estimating a robot coordinates in an external reference frame from sensor data, using a map of the environment* [Thrun et al., 2005, p. 3]. It *has been recognized as one of the most fundamental problems* for vehicles capable of locomotion [Fox et al., 1999]. Pose estimation (i.e., position, veloctity and attitude) is an essential part of the vehicle navigation. Navigation also requires knowledge about vehicle surroundings (obtained for instance by mapping or obstacle detection) and proper actions (e.g., route planning, decisions which area to explore) to achieve the goal of interest. However, the problem of knowing the pose of a vehicle is still challenging and localization, as the lower level of navigation, needs to be addressed to provide best possible performance.

In this thesis, we study local techniques to determine pose of a vehicle (e.g., a mobile robot) with respect to a local coordinate frame (i.e., incremental pose estimation, relative to the initial position). Therefore, initial position of the robot is required and localization system does not incorporate any absolute measurements in the global coordinate frame, such as GPS, or knowledge about the environment, such as maps or landmarks. Local techniques have obvious drawbacks. One of them is that the unobservable pose quantities have no certain error bounds, thus estimation suffers from uncompensated cumulative error. This is apparent mainly in position estimation. Moreover, the estimate cannot recover if the system loses track of the unobservable quantities, and implicitly cannot handle the well known kidnapped robot problem. Nevertheless, relative localization is generally vital subpart of any navigation system, even if the system implements global techniques exploiting global positioning services, knowledge or perception of the environment. Relative localization plays crucial step towards autonomous capabilities of a vehicle, since it ensures much needed estimation of the vehicle pose during unfavorable conditions. Such a situation may occur during unavailability of the absolute positioning (e.g., satellite navigation or landmark-based systems), missing or faulty map information, or any circumstances, that makes the updates with respect to the global coordination frame infeasible. One of many examples

is represented by the complex search and rescue missions, in which rescue teams deploy mobile robots to help them assess the situation (e.g., by exploring the disaster site and searching for victims). Robots deployed in these missions rarely have perfect operation conditions and struggle with harsh environment including static and dynamic obstacles, unreachable or underground areas, uneven terrain, ubiquitous dust, changing lightning conditions etc. All of these conditions influence the perception of the robot, make the navigation task difficult and in many instances the robot depends on the lowest level localization means it has—the local techniques. Therefore, we claim that the local techniques deserve attention as they contribute to the overall goal of having accurate, precise, and robust navigation system.

Although the pose and environment of the vehicle can be perceived through various sensors, the sensor observations are in many cases indirectly related to the system states. That means they are not observable directly and carry only partial information about the quantities of interest. Moreover, the sensory readings are usually corrupted by various noise terms. Therefore, the majority of vehicle state estimation used for vehicle navigation is based on the data fusion techniques (also referred to as sensor fusion). Broadly speaking, the goal of data fusion is to obtain more reliable and clear information in the world, which is not deterministic and includes models and sensing, which are not perfect. Techniques for data fusion are applicable in a wide range of areas and their popularity consists in the fact that data fusion provides in many cases improved and more accurate estimate, when compared to a single source of information [Hall and Llinas, 1997].

1.1 Motivation

First, we try to answer a fundamental question: What is the best way to implement a state estimation algorithm and which method should be used for a mobile robot. As this topic is rarely covered in the literature, we are motivated to seek reliable state estimation technique, which allows to incorporate relevant kinematic model, achieve sufficient performance and bearable computational load. There are many ways to determine the pose of the vehicle; however, Kalman filtering belongs to the most popular ones. The choice of the estimator often depends on the particular application. Kalman filtering excels in working with a uni-modal distribution, which is straightforwardly parametrized by its mean and covariance. On top of that, Kalman filters have moderate computational load when compared for instance to sampling approaches (e.g., particle filters, Monte Carlo methods). Therefore, Kalman filter is the algorithm of choice in this doctoral thesis.

Second, we are motivated by the general need for more accurate and reliable solution and bring the idea of extending the state estimation by various machine learning approaches. Our

goal is to provide improvement of the knowledge about the vehicle pose by the means of extracting additional information that may contribute to the overall localization performance. State estimation algorithms usually have predetermined conditions, which ensure that the estimator works properly and has characteristic behavior, such as stability of the estimate and convergence to the true value. However, the observed data in practice frequently break these conditions and cause inaccuracy or even failure of the estimator [Ndong and Salamatian, 2011]. Therefore, there is a need for analyzing the unwanted and undesired data. Such situations may occur in various scenarios and perception means: GPS readings in a urban area [Suzuki et al., 2011], or lost visual features in dynamically changing environments [Scaramuzza and Fraundorfer, 2011], [Kümmerle et al., 2013], to name a few. There is the possibility of observing patterns in the sensor data gathered in multiple scenarios and exploiting machine learning methods to extract useful information to aid the state estimation. Useful information, in the context of previously mentioned issues, can be refined by monitoring and rejection of the unwanted measurements. As we show later in this thesis, monitoring of the health of the Kalman filter observations is often overlooked; however, it should not be.

1.2 Structure of the thesis

This doctoral thesis takes the format of a *thesis by publication*, thereby it presents publications relevant to the topic of the thesis. This thesis format is approved by the Dean of Faculty of Electrical Engineering by the *Directive for dissertation theses defense, Article 1.*

The main contributions in this thesis are divided into Chapter 4 – 6. Chapter 4 and Chapter 5 present two publications with unified formatting. In these chapters, publications follow after a short summary sections, where we recapitulate the previous work, present the main topic, conclusions, and contributions of the work. Chapter 6 then extends the work in Chapter 5. All author's publications are listed in the publication list (see p. 90).

The thesis is organized as follows: Chapter 2 presents the current state-of-the-art in both state estimation as well as anomaly detection and approaches combining state estimation and machine learning. Contributions achieved are summarized in Chapter 3. Chapter 4 and 5 introduce author's two major publications related to the topic of this doctoral thesis. Chapter 6 is an extension of Chapter 5 and provides further examination of other relevant methods for anomaly detection. The doctoral thesis is summarized and concluded in Chapter 7, which also discusses the suggestions for future work.

Chapter 2

State of the Art

This chapter covers the state of the art regarding topics related to this thesis. In this thesis, we aim for a state estimation system capable of working in challenging environment, preferably for large scale scenarios during search and rescue missions. Therefore the first part of this chapter reviews the current state-of-the-art in state estimation applied especially for vehicle localization. This part reviews both ground and aerial localization and navigation systems, with the focus on deployment in large scale and challenging search and rescue missions. As the second part of the thesis aims for applying machine learning methods in combination with state estimation, we review two areas. The first area studies anomaly detection carried out by standard and machine learning methods. The second area introduces terrain recognition methods, which represent combination of state estimation and machine learning, which is similar to the topic studied in this thesis. Both these sections cover aiding and improving the state estimation framework by observing patterns and deducting effects of the environment on the navigation task.

2.1 State estimation problem

State estimation concerns the issue of using information obtained from the observed outputs to reduce the uncertainty about the system behavior [Maybeck, 1982, p. 1]. In the data fusion, the state estimation addresses the problem of determining the value of quantity of interest from sensor observations (also known as measurements or readings). There are many ways to deal with the uncertainty in the system; however, probabilistic methods are by far the most common technique appearing in engineering and applied robotics research [Durrant-Whyte, 2001, p. 1]. Probabilistic approaches are typically more robust to various aspects (e.g., sensor limitations, sensor noise, environment influence) and often scale well to complex and un-

structured environments. These properties make them very applicable to nearly every problem involving perception [Thrun et al., 2005, p. 3].

Certainly one of the most widely used estimation algorithms in present systems theory is the Kalman filter (KF) [Kalman, 1960]. It is a well established state estimation approach [Görner and Stelzer, 2013, Yi et al., 2007, Dissanayake et al., 2001] which excels in working with a unimodal distribution parametrized by mean and covariance. The Kalman filter is an estimator, which provides estimates of the state as well as its uncertainty [Bar-Shalom et al., 2001, pp. 381-394]. Majority of the real-world applications, including ground and aerial mobile robotics, is nonlinear and must be addressed by nonlinear techniques or approximations to maintain the performance and stability of the modeled system. Nonlinear problems are commonly handled by the Linearized KF (LKF), extended KF (EKF) or sample-based methods such as unscented KF (UKF) [Chen, 2003, Gustafsson et al., 2002]. Probably the most popular of the mentioned methods is the EKF, which has been applied in enormous number of applications where it achieved excellent performance [Grewal et al., 2007, p. 210]. Some of the essential EKF properties can be summarized as follows:

- the EKF linearizes the problem about the estimated trajectory [Grewal and Andrews, 2010] (as opposed to the Linearized KF, where the linearization of the system model is about the nominal trajectory; moreover, this trajectory has to be known in advance and does not depend on the measurement data [Grewal et al., 2007, p. 209]);
- it is defined as a suboptimal filter, since it applies the Taylor series approximations;
- it requires the derivation of the Jacobians;
- it is often applied in cases, where the model nonlinearities are not too high and do not cause instability [Vandyke et al., 2004].

Highly nonlinear problems are commonly handled better by the UKF [Julier and Uhlmann, 1997]). However, the computational complexity of the UKF is higher [Chen, 2003, Laviola, 2003]. That makes the UKF less feasible for running state estimation in real-time and on-board for large state space. The computational efficiency along with the need to approximate are one of the most cited fundamental limitations of probabilistic algorithms [Thrun et al., 2005, pp. 5-6]. With limited computational power, quite common factor in real-time processing in mobile robots, reasonable computational costs are one of the key factors of deploying a state estimation algorithm [Chen, 2003].

There are two standard ways of implementing the EKF: the total state space and the error state space configuration (also referred as the *direct* and *indirect* formulation, respectively) [Munguia and Grau, 2014]. The *direct* EKF is characterized by conceptual clarity—it

is straightforward to implement and setup the filter. This configuration incorporates a model, which is derived directly from the kinematics of the system. Despite the apparent advantages of the *direct* EKF configuration, it has several disadvantages over the *indirect* EKF [Roumeliotis et al., 1999]. Most importantly, the *direct* form is prone to temporary failures, which can cause the filter to fail and not recover. The *indirect* EKF is usually derived by linearizing the nonlinear differential equations using *perturbation analysis* [Shin, 2005]. Here the error state vector represents the difference between the estimate and the true value, and the covariance is related to the error state, not actual variables of interest. A review and tutorial to the readers for better understanding and correct implementing of the LKF and EKF in different configurations can be found in in [Jwo and Cho, 2010].

Essentially, there are multiple integration options in the data fusion. The integration architecture can be *loosely coupled* (the most computationally efficient approach, which handles the data separately) or *tightly coupled* (directly fuse the data sources and often achieve higher precision) [Li and Mourikis, 2013]. Depending on the fusion architecture, the filter works even in an open loop manner (feed forward) or closed loop manner (feedback) [Jwo and Cho, 2010]. In the end, the final Kalman filter implementation depends on the designer's choice. Often also the other aspects, such as computational power available or sensors and systems being fused, play crucial role in the decision.

2.2 State estimation for vehicle localization

The vehicle[1] state estimation is a very broad area. Thus, this section presents research related especially to unmanned ground or aerial vehicles. First, we present literature review of the general navigation of the ground and aerial vehicles. All these works employ platforms equipped with a common set of sensory modalities very similar to the sensory set used in this thesis. In the following review, modalities are represented by the proprioceptive[2] sensors, where the buildup of errors is inevitable, and exteroceptive[3] sensors, which are generally used to suppress or correct the build-up error. Therefore, the sensory set generally includes combinations of the inertial measurement unit (IMU), wheeled or tracked odometry (further referred to as odometry), Global Positioning System (GPS) receiver, a camera, a laser range finder; or other aiding sensors (e.g., magnetometers, distance sensors, inclinometers).

[1]The pose estimation of a vehicle is often studied by mobile robotics research groups, hence a vehicle can be also referred as a mobile robot (e.g., wheeled, tracked, or road vehicle, legged robot, fixed wing aircraft, helicopter, or multi-rotor helicopter.)

[2]Proprioceptive information is isolated from external aids and related to the measurements internal to the system

[3]Exteroceptive information is obtained through perceiving the surrounding environment

2.2.1 Ground vehicles localization

A popular relative[4] localization technique for a wheeled or tracked vehicle is the dead reckoning[5] using odometry. Wheels or tracks are equipped with encoders, which measure how far the wheels have rotated. This overall information about the motion of the vehicle is then used by the odometry, which estimates the velocities using a motion model. The odometry information then enables the vehicle to track its pose as a function of time [Siciliano and Khatib, 2008, p. 477]. When relying on odometry, the drift caused by wheel slippage, or insufficient odometry model generally leads to the buildup of errors over time [Cook, 2011, pp. 249-252]. Therefore, odometry is often combined with the IMU, which supports it with attitude estimates (i.e., orientation of the object) and may provide even full 3D relative pose estimation [Endo et al., 2007, Simanek et al., 2015]. Resulting localization system provides low-level proprioceptive-only localization and generally allows high sampling rate as well as processing rate (about 100 Hz [Yi et al., 2007]), usually without excessive computational load. In most cases it serves as a short term backup for navigation without any necessity of perceiving the surrounding environment. In the outdoor environment, the dynamically changing conditions often influence the exteroceptive modalities and cause degradation or complete unavailability of the information (obtained for instance by GPS, magnetometer, vision, range finding etc.). Hence, the necessity of having a reliable dead reckoning system and improving its performance is essential. However, there are drawbacks rising from the core principle of the dead reckoning and these have to be dealt with properly. The most common problems rest in the orientation and slippage errors, which can be remedied by numerous approaches. In [Yi et al., 2007] and [Anousaki and Kyriakopoulos, 2004], authors present a 4-wheel robot with an improved skid-steer model based on a Kalman filter estimating trajectory using velocity constraints and slip estimate. Alternative method appears in [Endo et al., 2007] where the IMU and odometry are used to improve tracked vehicle navigation via slippage estimates. A closed-loop control system approach for reducing the drift in the heading estimate is applied in [Borenstein and Ojeda, 2009]. Substantial effort has been also devoted to investigation of the odometry derived constraints [Dissanayake et al., 2001], or innovation of the motion models [Galben, 2011]. Concerning the works mentioned so far, localization approaches worked with dead reckoning realized in 2D. There also exist approaches providing 3D odometry, which is also referred to as 6 degree-of-freedom (DOF), derived from the rover-type multi-wheel vehicle design [Lamon and Siegwart, 2004], [Chugo et al., 2008]. All these works are closely related to our research, presented in Chapter 4, regarding the development of the EKF-based

[4] Also referred to as incremental or local; it requires the initial location of the vehicle and its application is limited due to buildup error; it usually cannot recover from serious localization errors, as opposed to global techniques, which can localize the object without any other prior knowledge.

[5] Derived from *deduced reckoning*; process of calculating the current pose of the vehicle based on the previous one and current measurements.

estimation architectures for a skid-steer mobile robot

Most of the 6DOF IMU and odometry dead reckoning solutions require other supporting sensor systems, especially exteroceptive, to assure more accurate, precise and reliable estimates used for autonomous navigation in a global sense. Authors in [Shen et al., 2011] demonstrate, that a very low-cost IMU and odometry dead reckoning system can be successfully combined with visual odometry (VO) [Scaramuzza and Fraundorfer, 2011], [Sakai et al., 2009] to produce more reliable navigation system. However, it is pointed out in [Rodriguez F et al., 2009], that a trade-off between precision and execution time has to be investigated in the approaches using the VO. Moreover, the VO degrades due to high rotational speed movements and it is susceptible to illumination changes and lack of sufficient scene texture [Scaramuzza and Fraundorfer, 2011]. Another typically used 6DOF aiding source is the laser scanner used for estimating vehicle motion by matching consecutive laser scans in a stop-and-go moving mode. In this manner, the laser scanner is capable of creating a 3D metric map of the environment [Yoshida et al., 2010], [Zhuang et al., 2011], [Suzuki et al., 2010]. One of the standard 3D odometry methods appears in [Morales et al., 2009] where it is used as the prediction step of the extended Kalman filter (EKF) and fused with the update step using GPS receiver, altimeter, and laser scanner road detection. It is also practically demonstrated that achieving reliable, robust and accurate state estimates is feasible through combining sufficient aiding sensors with the GPS. Rich sensory set used for full 3D pose estimation is presented in [Kubelka et al., 2015]. A tracked robot is equipped with IMU and velocity sensors (tracks odometry) for proprioceptive sensing, and omnidirectional camera and rotating laser range-finder to perceive the environment. Authors present a novel EKF framework and also employ experimental procedure revealing most frequent failure scenarios. These aspects form the multi-modal data fusion, which we extend in Chapter 5.

2.2.2 Aerial vehicles navigation

Nowadays, the aerial platforms deserve a lot of attention due to their unique ability of accessing otherwise unapproachable areas and complementing the information gathered from the ground. However, as stressed out by authors in [Bachrach et al., 2011], deploying an aerial platform rises many issues. These include mainly limited payload and flight time, fast dynamics, constant motion, omnipresent vibrations, indirect position estimates, as well as rapid and dangerous movement in unstructured or unknown environment, where a slight misstep may lead to the destruction of the platform. The most used aerial vehicles (also referred as drones, UAVs) in research have the shape of a small fixed-wing aircraft, or a multi-copter (also called multi-rotor) that are superior in their easier control and hovering capability. It is common, that a off-the-shelf civil multi-rotor for outdoor usage is equipped with an IMU and a GPS receiver. This setup together with the autopilot ensure attitude and position control, and also semi-autonomous maneuvers

such as take-off, way-point following, and landing. However, the main challenge comes, when the multi-rotor maneuvers, in a GPS-denied environment, such as dense urban areas or indoors. The GPS position information has to be replaced with other sensory system (e.g., laser range finder, camera), in order to maintain stable position and track the movement of the vehicle.

A typically equipped quad-rotor is presented in [Bachrach et al., 2011]. Authors employ the laser range finder, IMU, and a laser-reflecting mirror (to determine the altitude above the ground – it is assumed that quad-rotor operates over a flat floor). The on-board controller and IMU then stabilize the roll and pitch angles, position controller runs an EKF fusing the IMU and laser-based position increments, the planner and the SLAM algorithm provide delayed corrections to the EKF. With the described sensory set and algorithms, the quad rotor is able to fly and hover in constrained indoor and urban environments.

Representative of the aircraft type aerial vehicles is presented in [Bry et al., 2012] in a form of a fixed wing aircraft, and authors demonstrate that such a platform is able to fly in an indoor environment. The state estimation is based on an IMU and a planar laser range finder suitable for use in real-time on a fixed-wing air vehicle. Thus authors present a minimalist equipment allowing accurate state estimates during aggressive flight in unstructured 3D environments without the use of an external positioning system, however, given the 3D map of the environment represented as an occupancy grid. The pose estimates are determined by the EKF and a Gaussian Particle Filter, which is optimized to require much less particles than the standard approach. The accuracy of the state estimates are validated by Vicon[6] ground truth and also qualitatively by inspecting the reconstruction of the environment.

Despite the fact that the popularity of aerial vehicles deployment rises, there is a lot of open problems rising mainly from the nature of the 3D movement. More challenging environment also puts greater relevance of having correct sensor measurements for proper state estimation and following path planning and control. Such need inspired us in developing a framework that automatically rejects faulty measurements in a multi-modal data fusion framework presented in Chapter 5.

2.3 Challenging problems in vehicle navigation

Although the vehicle state estimation has been solved for many decades, there are still applications that are challenging for many reasons. In recent years, there has been much more pressure from the mobile robotics community to support the research with thorough experimental validation. The goal is to avoid limited duration of the experiments, which are generally conducted in

[6]Motion capture assembly of multiple infra-red cameras and reflective markers attached to a tracked object. Allows to capture accurate and precise 6DOF pose of the object.

relatively small, restricted, or simplified environments. At the same time, researchers are forced to carry out experiments in real environments and conditions. As the ultimate goal of autonomous navigation in natural environment is still not reached, there are research groups, which focus on these particular areas. In this thesis, we also aim to assess the experimental part by conducting large number of experiments in various environments, even those simulating the search and rescue missions. Therefore, this section puts the approaches presented in Section 2.2 into the context of large scale autonomy and deployment in search and rescue missions.

2.3.1 Large scale autonomy

The logical advance in vehicle navigation is to pay attention to the autonomy and understanding of large workspaces. Autonomous vehicles must interact safely with both the environment and users, move precisely in a dynamic environment and perform evenly under vast majority of conditions [Furgale et al., 2013]. The problem of the navigation a large scale dynamic environment is also mentioned in [McDonald et al., 2013]. Authors deal with the state estimation for combining multiple mapping missions that consists in aligning partial maps, which is closely related to multi-robot mapping problem. They combine successfully multiple visual SLAM missions performed repeatedly over time in the same environment. Authors set the goal, which consists in tens to hundreds of repeated missions, and the experiments are conducted in different indoor and outdoor sessions. This work is a great example of the comprehensive quantitative assessment and confronting the problem of perceiving large environment through a different approach combining multiple similar experiments.

Authors in [Konolige et al., 2011] show that just integrating the IMU in a loosely coupled structure with VO can dramatically improve the long term accuracy of the VO. Further improvements of the VO, such as novel scale-space features, sparse bundle adjustment, can upgrade the IMU/VO fusion to an alternative to the standard IMU/GPS approaches to car navigation. Moreover, the error statistics of the described IMU/VO fusion perform better than the standard GPS-based fusion for over distance up to 10 km on a vehicle running up to 5 m/s.

Dynamic changes play a crucial role also when navigating in urban environment. This problem is addressed by authors in [Kümmerle et al., 2013] where they propose a navigation system for pedestrian-like autonomous robot maneuvering in a city environment. They utilize a set of laser range finders, an IMU, a GPS receiver, a priori known map, and variant of the Particle Filter to perform SLAM in a city with crowded streets. In such surroundings, they have to deal with the vegetation detection, dynamic obstacles tracking, 3D obstacles detection, or vibration-based ground classification to properly navigate the robot. In the end, they demonstrate the performance in a 3 km long experiment, where the robot successfully moves through the city. For the given goal, there remain unsolved problems partially originating from the platform de-

sign constraints (e.g., running over uneven terrain), but mainly from the dynamically changing environment (e.g., fast moving objects, objects falsely perceived as not traversable).

Partial capability of autonomous behavior is also demonstrated by authors in [Chowdhary et al., 2013]. For a GPS-denied navigation and real-time closed-loop control, they implement an EKF-based fusion of IMU, monocular VO, and a sonar. The autonomy includes especially landing and take off maneuvers of multiple types of the UAVs in both outdoor and indoor environments, and real-world conditions. To perform well in longer experiments, they implement a novel algorithm to manage a database of visual features based on the confidence in each measurement. These results confirm the sustainability of using the IMU/VO for autonomous GPS-denied environments.

A 6-DOF pose estimation deployed on a quadruped system combining the stereo-camera, IMU, odometry, and optional intermittent GPS position updates within the EKF is capable of reaching 1 % position error of distance traveled [Ma et al., 2015]. Authors in this work report, that this specific version of the EKF is the most tested of any quadruped systems with up to 32 hours of robot operation and over 124 km of robot travel. As this type of robot was developed to carry load, such endurance and localization performance is particularly important for autonomous following a human leader. According to our opinion, this is by far the most elaborate experimental evaluation we have encountered in the field of mobile robotics.

2.3.2 Search and rescue

Search and rescue is another area where the localization is a crucial step towards autonomous behavior. As the mobile robots are deployed more and more frequently in such missions, the focus is maintained especially on cooperation with professional rescue teams, usually firefighters. Research concerning area of search and rescue focuses on partial autonomy, which helps to lower the workload of the tele-operation. Lowering the operator burden can be represented by implementing semi-autonomous behavior. Such improvements are reported for instance in [Okada et al., 2011], where the flippers[7] of the robot are controlled autonomously using 3D scanning; or in [Ellekilde et al., 2007], where the 3D scanning ensures a clear and understandable information given to the end user.

As the disaster site during a search and rescue mission is usually not safe for any platform, researchers aim for multi-robot cooperation to mitigate the disadvantages of a single-platform deployment. For instance, authors in [Michael et al., 2012] present a team of ground and aerial robots. One UGV with sensor suite allowing generation of the 3D maps, second UGV with a helipad and a quad-rotor on-board. The aerial robot is used to explore areas otherwise unreachable by the UGVs and performed only when landing and take-off is possible. The goal of such coop-

[7] One or two pairs of sub-tracks allowing the ground robot obstacle traversing.

eration is to generate 3D multi-floor maps to analyze if there are any structural reinforcements needed.

Similar 3D mapping using a tracked UGV is presented by authors in [Nagatani et al., 2011]. They also define a set of milestones for a multi-robot cooperation: autonomous traversal of uneven terrain, development of a system for the continuous acquisition of the 3D data of the environment, coverage path planning, centralization of map data obtained by multiple robots, and fusion of map data obtained by multiple robots.

A method for robust and efficient loop closure in harsh large-scale search and rescue environments is demonstrated in [Kleiner and Dornhege, 2007] and utilizes deployed radio-frequency identification makers for landmark data association and slippage-sensitive odometry for 2D pose tracking. Despite the fact, that the markers have to be deployed before the mission, they present computationally efficient method for building elevation maps. This method utilizes an EKF for 3D pose tracking based on laser scan matching supported by the VO and can be applied in real-time while navigating on rough terrain.

Table 2.1: State estimation reference list. Provides the summary of related work and used algorithms, sensors and platforms.

State estimation algorithms	
EKF	[Konolige et al., 2011] [Bachrach et al., 2011] [Ma et al., 2015] [Bry et al., 2012] [Kleiner and Dornhege, 2007] [Yi et al., 2007] [Anousaki and Kyriakopoulos, 2004] [Dissanayake et al., 2001] [Galben, 2011] [Morales et al., 2009]
UKF	[Galben, 2011] [Sakai et al., 2009]
PF	[Kümmerle et al., 2013] [Bry et al., 2012] [Yoshida et al., 2010] [Suzuki et al., 2010]
Application	
GPS-denied	[McDonald et al., 2013] [Konolige et al., 2011] [Chowdhary et al., 2013] [Bachrach et al., 2011] [Kubelka et al., 2015]
Long-term autonomy	[McDonald et al., 2013] [Konolige et al., 2011] [Michael et al., 2012] [Ma et al., 2015]
Search and rescue	[Michael et al., 2012] [Nagatani et al., 2011] [Okada et al., 2011] [Ellekilde et al., 2007] [Kleiner and Dornhege, 2007] [Kubelka et al., 2015]
SLAM	[McDonald et al., 2013] [Kümmerle et al., 2013] [Bachrach et al., 2011] [Michael et al., 2012] [Nagatani et al., 2011] [Ellekilde et al., 2007] [Kleiner and Dornhege, 2007]
Sensory set	
IMU/GPS	[Dissanayake et al., 2001]
IMU/ODO	[Michael et al., 2012] [Kleiner and Dornhege, 2007] [Yi et al., 2007] [Anousaki and Kyriakopoulos, 2004] [Endo et al., 2007] [Borenstein and Ojeda, 2009] [Lamon and Siegwart, 2004] [Chugo et al., 2008]
IMU/ODO/VO	[Shen et al., 2011] [Sakai et al., 2009]
IMU/ODO/VO/GPS	[Ma et al., 2015]
IMU/VO	[Konolige et al., 2011] [Chowdhary et al., 2013]
IMU/Laser	[Kümmerle et al., 2013] [Michael et al., 2012]
IMU/ODO/Laser	[Yoshida et al., 2010] [Suzuki et al., 2010]
IMU/VO/Laser	[Bachrach et al., 2011]
MU/ODO/VO/Laser	[Kubelka et al., 2015]
IMU/Laser	[Bry et al., 2012]
ODO/Laser	[Nagatani et al., 2011] [Okada et al., 2011] [Zhuang et al., 2011]
VO	[McDonald et al., 2013] [Scaramuzza and Fraundorfer, 2011] [Rodriguez F et al., 2009]
VO/Laser	[Ellekilde et al., 2007]
Platform	
Ground	[McDonald et al., 2013] [Konolige et al., 2011] [Kümmerle et al., 2013] [Yi et al., 2007] [Ma et al., 2015] [Anousaki and Kyriakopoulos, 2004] [Endo et al., 2007] [Borenstein and Ojeda, 2009] [Dissanayake et al., 2001] [Galben, 2011] [Lamon and Siegwart, 2004] [Chugo et al., 2008] [Rodriguez F et al., 2009] [Yoshida et al., 2010] [Zhuang et al., 2011] [Kubelka et al., 2015]
Aerial	[Chowdhary et al., 2013] [Bachrach et al., 2011] [Bry et al., 2012] [Shen et al., 2011]
Collaborative	[Michael et al., 2012] [Nagatani et al., 2011]

2.4 Fault detection in mobile robotics

The field of robotics and autonomous systems often addresses the problem of failure identification by fault detection and isolation (FDI) methods [Pettersson, 2005]. Goal of the FDI is to maintain the ability to continue functioning after a sensor, actuator or mechanical failure has occurred—one of the most important problems in development of reliable and robust robots [Roumeliotis et al., 1998]. According to [Gertler, 1998, p. 3] fault is defined as an anomaly in behavior of the monitored system and can be detected (indication that something is going as not expected), isolated (locating the faulty component) and identified (determination of magnitude of the fault).

Mobile robots used in real-world situations must deal with many uncertainties. Generally, there are various ways in the robotics community to address the FDI problem:

- quality inspection of the provided information [Brunner et al., 2013],
- equipping platforms with redundant information sources [Sundvall and Jensfelt, 2006, Mendoza et al., 2012],
- monitoring the information flow between the control and actuation [Christensen et al., 2008],
- monitoring the reliability of resources [Morales et al., 2008],
- proper recognition and modeling of the sensor and mechanical failures [Goel et al., 2000].

In the following sections, we focus mainly on anomaly detection, which can be specified as monitoring of the sensory information and proper recognition of sensor failures.

2.4.1 Anomaly detection

Anomaly detection, a subpart of the FDI, addresses the task of finding patterns that do not conform to expected behavior [Chandola et al., 2009]. There has been a lot of effort put into identification of such disturbances in a broad range of different applications, e.g., network security, fraud alert, medical domain, etc. [Dua and Du, 2011].

For the purpose of differentiating fault detection approaches in this thesis, we further divide the anomaly detection into implementations using standard and machine learning techniques. Standard FDI techniques implementing anomaly detection can be divided the into following groups [Hwang et al., 2010]:

- residual generation (observer methods, Kalman filter, nonlinear systems, etc.),
- decision tools (such as cumulative sum algorithm, generalized likelihood ratio test)

- reconfiguration (multiple or adaptive model approach).

The machine learning techniques are mostly dependent on previously gathered data and allow the development of anomaly detection methods, that are non-parametric, or even adaptive to the changes in the monitored system [Ahmed et al., 2007]. The goal of machine learning is defined as developing an algorithm that learns to perform specific task from the past experience and this task can be handled in supervised, unsupervised, semi-supervised or reinforcement manner [Bishop and Nasrabadi, 2006]. Approaches utilizing the supervised learning are usually provided with an training set[8] and the algorithm tries to learn the hypothesis function[9]. The unsupervised approach is given only the raw data and tries to find the hidden structure within and represent the data in more informative way. Finally, in the reinforcement learning, the algorithm receives a feedback that gives a reward after some action. Therefore, the focus of this approach is aimed to the on-line functionality.

The availability of previously seen measurements can be exploited for improving the state estimates, reliability, robustness, or for gaining crucial information related directly to the state estimation problem. Since we build on our expert-labeled datasets of UGV experiments and exploit both standard and machine learning methods, following section presents works related especially to the standard statistical tests, and supervised and unsupervised machine learning applied in the field of anomaly detection.

2.4.2 Methods for anomaly detection

Methods for anomaly detection can utilize both statistical tests and machine learning techniques. In this section, we provide a review of the commonly used methods from both areas with the application in the multi-sensor data fusion.

A typical application of the statistical tests is presented in [Soule et al., 2005] with the focus on supervising large-scale network traffic. Authors study the network traffic patterns by anomaly detection based on Kalman filter residuals. They examine four different threshold based schemes: the instantaneous residual traffic to a threshold, inspection of the residual mean (using cumulative sum algorithm and generalized likelihood ratio test), changes in variance estimated by the Kalman filter, and multi-scale variance shift using a wavelet transformation approach. Authors also comment on the drawbacks of their validation methodology – labeling the time line by visual inspection and synthetic validation data sets. According to their evaluation, the first approach can introduce errors such as missing anomaly or generating false alarms in the detection. Despite the second approach may provide perfect control over the data sets, it may not cover the real-world behavior. The difference between the real and synthetic data set is also apparent in

[8] A labeled set consisting of input and output values.
[9] A mapping from the input to the output value.

the comparison of the analyzed methods. Authors report slight inconsistency in the performance when evaluating methods on both data sets, meaning that the synthetic data set does not exactly mirror the real-world problem. Despite the fact that the method detecting changes in the mean estimates introduce a lag before reaching a decision, it performs best in the final comparison.

Another case of using statistical decision in a multi-sensor data fusion is demonstrated in [Caron et al., 2006]. The aim of the work is to develop a GPS/IMU fusion algorithm with a possibility of erroneous data rejection. In order to have a supervised fusion process, authors define a contextual space for sensor validation and use fuzzy logic to determine the reliability of the observed GPS data. The validity domains are based on the standard Mahalanobis distance[10] q_i and a χ^2 distribution test, checking the distribution of sum of squares of the independent normally distributed random variables. These tests assume unusable sensor readings, if the value q_i is beyond a threshold defined by a 95 % confidence level of the χ^2 distribution. Authors also suggest that the work should be investigated further for subset testing allowing more observations to be used to remove erroneous measurements.

Mahalanobis distance is a popular choice in the field of Kalman filtering applied in robotics [Bloesch et al., 2013]. It provides more robust version of the filter just by thresholding the innovations. Moreover, if it is appropriately chosen (e.g., hand-tuned for a particular data set), it leads to near-optimal filtering [Ting et al., 2007].

The problem of dealing with non-Gaussian distribution of the innovations in the Kalman filter is addressed in [Ndong and Salamatian, 2011]. Authors state that in practice, the innovation process cannot be expected to be white and Gaussian, since the observed data themselves frequently break this property. Such a statement is based on the results, which repeatedly indicated a large number of false alarm detections, when assuming the process to be Gaussian and white. Therefore, authors assume that the real distribution of the residuals is an ensemble of Gaussians (i.e., a Gaussian Mixture Model (GMM) estimated by the Expectation Maximization iterative algorithm), and that there is a time dependency captured by a Hidden Markov Model (HMM). With this approach they achieve decrease of the false alarm rates. The residual sequence is distributed into k clusters using the GMM method and afterward classified by the HMM into several states, each being formed with part of the k families, to find the temporal dependencies above the components.

Similar issue of detecting abnormal behavior via combination of several measurements in the Kalman filter is reported in [Knorn and Leith, 2008]. In addition to a self-tuning Kalman filter estimating the CPU usage, authors incorporate a moving average for straightforward smoothing of the log likelihoods of the past measurements. The results indicate that such a standard signal processing yields significantly more robust anomaly indicator than a decision based only on a

[10] A scale-invariant distance that takes the variance of the variable into account.

single time instance. However, the expertise in setting a threshold is still required and based on data series inspection.

A method for finding a threshold for anomalous pattern rejection is proposed in [Sagha et al., 2013]. Generally, it consists of creating an anomaly free training set and learning the threshold automatically as the maximum value of a particular indicator. This represents the one-class classification on imbalanced data, where outliers or anomalous patterns are discarded, thus no longer distracts the training phase. Authors present a decision process in an ensemble of classifiers where each of them detects undesired observations in the data sets. Classifiers combined together increase the robustness of the entire system. Given two techniques, distance and information based, they decide if the classifier is healthy or anomalous. The first approach uses the averaged Mahalanobis distance between each classifier and the final fusion output, and compares the difference with a predefined threshold. The second approach tracks changes in the mutual information[11] of a classifier with the rest of the classifiers. These two methods are compared to the generalized likelihood ratio (GLR) test and one-class Support Vector Machine (SVM). As the authors comment, the one-class SVM is very sensitive to the chosen parameters[12], which are set experimentally according to the detection accuracy. Authors state, that their approach is comparable to the GLR and SVM, and does not need to monitor the raw data, as it only processes the decisions made by the classifiers.

A modified SVM classifier is presented in [Amer et al., 2013] to address the problem of one-class SVM sensitivity to the parameters and outliers in unsupervised learning. There are following steps that contribute to the robustness of the classifier: the slack variables[13] are modified to shift the boundary towards normal points, thus outliers are no longer support vectors; explicit outlier suppression mechanism (using additional outlier parameter variable, that makes sure the outliers do not contribute to the optimization objective) is introduced. Authors compare their solution to the standard one-class SVM, nearest neighbor method, histogram based outlier score, and clustering algorithms. In the comparison, the modified one-class SVM performs better than cluster and histogram-based outlier factors. Despite the fact that SVM parameter tuning raises the time complexity, it is shown that SVM based algorithms, especially one-class version, can perform reasonably well for unsupervised outlier detection.

[11] A method, which quantifies the shared information between two random variables.

[12] The upper bound of the fraction of the training data considered as outside of the hyperplane, and the bandwidth which affects the smoothness of the decision boundary.

[13] A measure in soft margin SVM, which specifies the degree of misclassification

Table 2.2: Anomaly detection reference list. Provides the summary of related work and used algorithms.

Rule-based	[Soule et al., 2005] [Caron et al., 2006]
Mahalanobis distance	[Caron et al., 2006] [Sagha et al., 2013] [Ting et al., 2007] [Bloesch et al., 2013]
χ^2-distribution	[Sagha et al., 2013] [Caron et al., 2006]
Log-likelihood ratio	[Sagha et al., 2013] [Knorn and Leith, 2008]
Information theory	[Sagha et al., 2013]
Linear/Logistic regression	
Neural networks (NN)	[Goel et al., 2000]
Support Vector Machine (SVM)	[Amer et al., 2013]
k-Nearest Neighbor (k-NN)	[Amer et al., 2013]
Kernel Density Estimator (KDE)	[Laxhammar et al., 2009]
One-class SVM (OCSVM)	[Amer et al., 2013] [Sagha et al., 2013]
Principal Component Analysis (PCA)	[Ndong and Salamatian, 2011]
Gaussian Mixture Models (GMM)	[Ndong and Salamatian, 2011] [Laxhammar et al., 2009]

2.5 Terrain recognition for ground vehicles

We introduce the problem of the robotic terrain classification (RTC) [Tick et al., 2012] as it often represents the combination of state estimation and machine learning methods—very similar to what we pursue in this thesis. For a ground vehicle, terrain recognition is an important part of the autonomous behavior [Tick et al., 2012]. The most common problem for ground vehicles using odometry is the drift of the trajectory estimates caused by slip and skid events occurring during the motion of the vehicle on harsh terrain (e.g., sand, soil, gravel, etc.). This drawback can be partially remedied by extending the sensory set by complementary modalities, either incremental such as visual and range-finding, or absolute positioning such as GPS and fixed beacon-based system. However, if the vehicle is able to detect these problematic areas, it can consequently avoid them, reduce speed, or modify the navigation strategy with the overall goal of improving the belief about its state.

Most of the RTC approaches are based on features extracted from either visual or inertial data and discriminating the terrain according to a training set. As the visual data processing is often computationally demanding and in many cases used for other purposes (e.g., visual odometry, object tracking), the vision-based RTC represents the smaller part of deployed solutions. A typical approach is given in [Howard and Seraji, 2000], where the stereo camera is used together with fuzzy logic to detect the terrain roughness (i.e., smooth, rough, rocky) and slope (i.e., flat, sloped, step). Final actions control is based on traversability labels and navigation rules. There are also approaches combining both visual and inertial data. In [Weiss et al., 2008] authors use the SVM

on the texture-based histogram features to classify the terrain in front of a mobile robot into 14 classes. The vibration data from accelerometers are used to verify the former prediction with the use of Fast Fourier Transform (FFT) based features. Similar method of combined classifiers is shown in [Halatci et al., 2007], where the Bayesian Classifier fusion is applied to two GMM and SVM processing the color and texture of the image, and vibration propagating through a rover wheel for a planetary exploration rover. Authors report average accuracy of 84 % for the fused classifiers, which is an improvement of 8 % over vibration and image only.

When it comes to the features derived solely from inertial data, most authors use the accelerometers, which provide natural information about the surface irregularities when traversing them. The dependency of the classification performance on the speed of the vehicle is pointed out in [Weiss et al., 2006] along with the fact, that suspension mechanism or soft wheels dampen the vibrations and the distinctiveness decreases. Authors use the SVM and features derived from power spectral density, FFT, and time domain, resulting in total misclassification rate of 5 %. Their work continues in [Weiss et al., 2007], where the SVM is proved to work well in comparison with neural networks, naive Bayes, decision trees and k-nearest neighbor for an outdoor robot driving between 0.2 and 0.6 m/s. The gained knowledge about the surface type is then utilized in [Komma et al., 2009], where authors extend the previous work to rely not only on one single classification, but on several consecutive observations. They apply a Bayes filter that improves the prediction performance by involving the classification history. A problem of feature selection is addressed in [Tick et al., 2012] and authors use an audio toolbox to extract more than 800 features from acceleration and angular rate measurements. These are narrowed down by sequential feature selection to a set of 15 to 20 most useful ones, that are utilized to train a linear Bayes classifier. A popular technique, neural networks, is evaluated for terrain recognition using IMU measurements in [Jitpakdee and Maneewarn, 2008] and reported to work with 80 % and 100 % accuracy when classifying the surface into 5 categories (flat, rugged, incline, grassy, and unclassified). Neural networks are also utilized in [Giguere and Dudek, 2011] in an unusual setup that consists of a metallic rod with a single-axis accelerometer, which is passively dragged along the surface. Authors compare the results with various time domain features (mean, variance, skewness, kurtosis, fifth moment, different threshold crossings, sum of variation over time, sum of higher half of amplitude spectrum) for both supervised (90 % accuracy for a 1 s window) and unsupervised neural networks (75 % accuracy for a 1 s window). Another frequently used technique, the principal component analysis (PCA), is studied in [Brooks and Iagnemma, 2005, DuPont et al., 2008] and used for feature extraction and especially for the dimension reduction.

The ubiquitous speed dependency of vibration features, which is rarely discussed and often solved as setting the speed constant, is commented in [Coyle et al., 2011, Collins and Coyle, 2008]. The attention is paid to the non-constant speed distribution that requires collecting large data sets for algorithm training. Contribution of motor action and different sensory modalities to

discriminative ability of a RTC system applied in a legged robot is discussed in [Hoffmann et al., 2014]. It is pointed out that the gait regime of the robot (i.e., pattern of movement) should not be neglected and rather form gait-sensor feature combinations to improve the RTC discrimination performance.

Recent development in RTC aims for self-learning frameworks [Santamaria-Navarro et al., 2015], where the terrain model is progressively updated on-line Combining different sensor modalities brings advantages also in the RTC area. Especially popular choices are for instance based on the radar data (using methods such as GMM and Mahalanobis distance) or monocular vision. The vision is in this work supervised by a radar classifier and trained in a modified color space more robust to changes in the light intensity.

Combination of a low-level and high-level RTC appears in [Milella et al., 2015] with the goal of generating traversability maps for the computation of optimal paths. Full 3D RTC in outdoor environments is divided into two modules: low-level heuristic detection of non-traversable areas (obstacles and holes) in front of the robot using a time-of-flight sensor; and a module using 3D LIDAR-based point cloud for high-level supervised classification in a cloud map. Authors exploit Gaussian processes (both classification and regression), trained with either positive samples only, or hand-labeled data. With only two features (local roughness, slope) the Gaussian process models achieve better performance than naive parametrization or SVM implementations with different kernels.

Authors in [Reinstein et al., 2013] contrary to standard RTC aim for direct terrain-based correction of the odometry estimates of a skid-steer vehicle. The work uses the linear regression to predict a coefficient, which is derived from the odometry and ground truth data, and defined as a scaling factor correcting the position estimates. This way, the trained regression successfully reflects the influence of the different types of surface on the performance of the odometry, instead of just perceiving the type of the terrain.

Table 2.3: Terrain classification reference list. Provides the summary of related work and used algorithms.

Rule-based	[Howard and Seraji, 2000]
Mahalanobis distance	[Santamaria-Navarro et al., 2015]
χ^2-distribution	[Santamaria-Navarro et al., 2015]
naive-Bayes	[Hoffmann et al., 2014] [Milella et al., 2015]
Linear/Logistic regression	[Reinstein et al., 2013]
Neural networks (NN)	[Weiss et al., 2007] [Jitpakdee and Maneewarn, 2008] [Giguere and Dudek, 2011]
Support Vector Machine (SVM)	[Weiss et al., 2008] [Halatci et al., 2007] [Weiss et al., 2006] [Weiss et al., 2007] [Komma et al., 2009] [Hoffmann et al., 2014] [Milella et al., 2015]
Gaussian Processes (GP)	[Milella et al., 2015]
k-Nearest Neighbor (k-NN)	[Weiss et al., 2007] [Komma et al., 2009]
Singular Value Decomposition	[Coyle et al., 2011]
Principal Component Analysis (PCA)	[Brooks and Iagnemma, 2005] [DuPont et al., 2008]
Gaussian Mixture Models (GMM)	[Halatci et al., 2007], [Santamaria-Navarro et al., 2015]

Chapter 3

Contributions

This dissertation investigates methods for pose estimation of a skid-steer mobile robot intended for search and rescue operations. The main goal is to provide a framework for relative localization of a multi-modal (i.e., multi-sensor) system and improve the framework with machine learning methods to obtain more accurate and reliable outcome. This thesis presents following contributions:

1. **Development of data fusion for mobile robot state estimation.**

   In order to represent the state of a mobile robot, we implement the relative localization, the dead-reckoning The objective is to find a reliable state estimation technique for observing the state of a vehicle (e.g., position, velocity, and attitude). Since different approaches to the architecture of a state estimator lead to different performance and computational demands, we seek the best possible solution and deploy the extended Kalman filter in various forms.

2. **Development of anomaly detection framework for the multi-modal data fusion framework.**

   Incremental pose estimation is prone to measurement errors, because they accumulate in time. General idea is to overcome the errors or disadvantages by deploying multiple sensory modalities and exploiting their complementary properties. In pursuing the goal of the best possible dead-reckoning, the measurements have to be monitored and anomalous observations, which spoil the pose estimate, have to be handled properly. Thus, we develop an anomaly detection system supporting the multi-modal state estimation framework. We apply the standard techniques intended for monitoring of the Kalman filter observations and go beyond by exploring different state-of-the art machine learning methods.

3. **Testing of the frameworks on real platform and in real-world scenarios**

   Last contribution of this thesis is related to both previously stated aims—thorough testing on challenging datasets and in real-world scenarios. We feel that this step is a fundamental part of every application involving a mobile robot. Therefore, we verify and evaluate the proposed algorithms by the means of real-world, preferably long-term experiments in a challenging environment. All of the algorithms are evaluated with the data gathered by a real platform—a skid-steer mobile robot intended for search and rescue missions. We compare the results quantitatively against ground truth data (if available). The verification datasets include several kilometers of distance driven, and are carried out both in indoor and outdoor environments.

Chapter 4

EKF-based Estimation Architectures for Data Fusion

4.1 Summary of the contributions

This chapter contributes mainly to the first contribution of the thesis (see Chapter 3). We analyze and evaluate different modeling approaches based on the direct and indirect form of the extended Kalman filter, well proven and reliable technique. We propose architectures combining the EKF with a complementary filter (CF), a state estimation technique, which was introduced in [Kubelka and Reinstein, 2012] for attitude estimation. We build these comparisons on the experience gained in the previous work applying Kalman filtering in a legged robot localization [Reinstein and Hoffmann, 2011, Reinstein and Hoffmann, 2013].

Total of four EKF configurations are examined for performance, computational burden, observability and consistency, and are deployed in a search and rescue mobile robot equipped with inertial measurement unit and velocity encoders. This way we combine two proprioceptive modalities, the IMU and the odometry (velocity encoders), which create a dead-reckoning localization able to localize the mobile robot with up to a certain performance limit. This performance limit is evaluated experimentally and is estimated to be around 4% of the final position error of the total distance driven (e.g., average of 4 m final position error for a 100 m trajectory). We also address the third contribution of the thesis by performing the evaluation on datasets gathered in both indoor and outdoor experiments with more than 4 km of distance traveled.

We show that despite the proposed EKF-CF hybrid architectures are sub-optimal and do not provide uncertainties estimates for all states, they are reliable, light-weight and sufficient dead-reckoning implementation of the pose estimation. This chapter focuses on the following aspects and brings the following contributions:

- correction of the odometry and overall transformation of the odometry-based localization from planar 2D to full 3D solution (6-DOF pose estimation)
- deployment of the stable attitude estimates from the complementary filter in dead-reckoning localization
- enhancing the standard IMU/ODO models by non-holonomic constraints
- lowering of the computational burden by combining estimation techniques
- thorough experimental evaluation
- implementation of selected EKF algorithm for on-board processing in NIFTi project[1] and deployment in rescue missions consisting of reconnaissance of earthquake-affected areas[2]

4.2 Publication

The work is represented by a publication with modified formatting and follows on the next page.

[1] EU FP7 NIFTi. "NIFTi project summary" YouTube, Jul 28, 2015. https://youtu.be/zygydwIE6s4

[2] For more details see the European Commission Memo, Aug 18, 2015, http://europa.eu/rapid/press-release_MEMO-12-620_en.htm

IEEE/ASME TRANSACTIONS ON MECHATRONICS, VOL. 20, NO. 2, APRIL 2015
The final publication is available at IEEE via http://dx.doi.org/10.1109/TMECH.2014.2311416

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

Jakub Simanek · Michal Reinstein · Vladimir Kubelka

Accepted: 27 February 2014 / Issue date: April 2015

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

Keywords urban search and rescue · complementary filter · extendedKalman filter

J. Simanek
Department of Measurement, Faculty of Electrical Engineering, Czech Technical University in Prague, Technicka 2, 16627 Prague, Czech Republic
E-mail: simanjak@fel.cvut.cz

M. Reinstein
Department of Cybernetics, Faculty of Electrical Engineering, Czech Technical University in Prague, Karlovo namesti 13, 12135 Prague, Czech Republic
E-mail: reinstein.michal@fel.cvut.cz

V. Kubelka
Department of Cybernetics, Faculty of Electrical Engineering, Czech Technical University in Prague, Karlovo namesti 13, 12135 Prague, Czech Republic,
E-mail: kubelvla@fel.cvut.cz

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

1 Introduction

The recent progress in sensor technologies and increase in onboard computational power brings new demands and pushes the limits of navigation of autonomous robots. Mobile robots are becoming increasingly more reliable and hence more popular even for complex missions such as Urban Search and Rescue (USAR). In general, there exist a number of different solutions to data fusion for localization and navigation, making it often unclear when deciding for one that ensures optimal performance. Moreover, the performance is influenced not only by the choice of the platform, its morphology and the sensor suite, but also by environment and mission specifications. Development of the system for USAR missions aims primarily on human-robot teaming, especially deploying an unmanned ground vehicle (UGV) (see skid-steer robot UGV in Fig 1) in a close cooperation with professional rescue teams. The mission, in which the testing UGV was deployed and used one of the EKF architectures proposed in this paper, consisted of reconnaissance of earthquake-affected areas[2](refer to (Kruijff, G.-J. et al. 2014) and (Kruijff et al 2012) for details about the USAR challenge and mission, respectively).

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

In order to ensure reliable performance, explore the disaster site, and perform high quality 3D mapping of the environment (such as in (Mobedi and Nejat 2012)) reliable dead reckoning development was essential. Since the extended Kalman filter (EKF) is a well proven and commonly used state estimation technique (Ma et al 2012; Arsenault et al 2011), the motivation for this paper lied mainly in seeking the most suitable EKF architecture. To our best knowledge, we are convinced no such comparison of estimation architectures covering EKF-based dead reckoning was carried out for mobile skid-steer robots intended for real USAR missions. Thus, the results of the analysis might prove beneficial to anyone asking the fundamental question: What is actually the best way to implement the EKF? Therefore, objective of this paper lies in comparing four different approaches: a nonlinear model (Nemra and Aouf 2010) and an error model (Shin 2005), each with and without acomplementary filter (CF) for attitude estimation (Kubelka and Reinstein 2012; Vasconcelos et al 2011). The performance of attitude estimation using the CF was thoroughly evaluated as part of our previous work in (Kubelka and Reinstein 2012), including testing of various filters to cope with inertial signals strongly affected by vibration. In this work, dead reckoning is represented by estimation of the six degree-of-freedom (6DOF) pose of the UGV, using the proprioceptive sensors only: odometry obtained from motor encoders and an inertial measurement unit (IMU) that consists of accelerometers and gyroscopes providing specific force and angular rate measurements (Titterton and Weston 1997, p. 26). As it was shown repeatedly, the combination of IMU and odometry is a popular technique to localize a mobile robot even in case of dynamic legged robots (Reinstein and Hoffmann 2011). It generally allows high sampling and processing rate, usually without excessive computational load. However, there are drawbacks, such as inertial navigation drift and wheel slip (Cook 2011, p. 249), which reflect also to the principle of dead reckoning. Therefore, substantial effort has been made to investigate improvements through slip estimation, velocity constraints (Yi et al 2007; Endo et al 2007), odometry derived constraints (Dissanayake et al 2001), or innovative motion models (Galben 2011). We have also addressed the slip compensation problem in (Reinstein et al 2013); however, our aim is not to develop a dead reckoning with bounded position error, which is in principle not achievable using proprioceptive sensors only (Aghili and Salerno 2013). Therefore, most of the state-of-the-art 6DOF IMU and odometry dead reckoning solutions require other supporting sensor systems, especially exteroceptive, to assure desired precision and reliability given by the target application; for example see (Aghili and Salerno 2013; Song et al 2011; Shen et al 2011; Morales et al 2009). The contributions of this paper are: first, providing the reader with analysis regarding the performance of different architectures for state estimation using the EKF; second, extending our previous work about the CF for attitude estimation (Kubelka and Reinstein 2012) to full 6DOF dead reckoning; third, testing in both indoor and outdoor structured environment (including stairs and other obstacles); and fourth, implementing the best approach for onboard processing in the Robot Operating System (ROS) (WillowGarage 2012b). This paper is structured as follows: Section 2 covers details regarding the theory and methodology. Section 3 presents experimental results and performance evaluation. Section 4 provides implementation details. Conclusions are given in Section . All the symbols used can be found in Table 1.

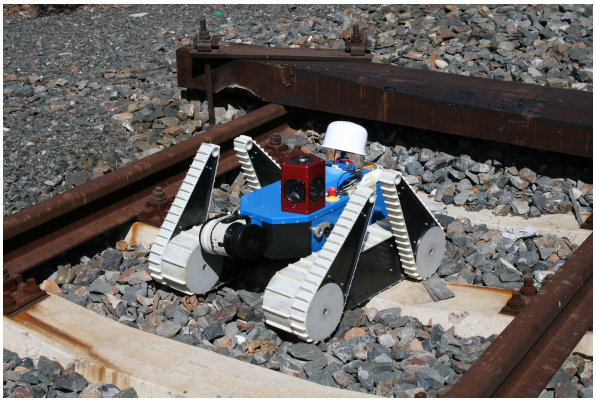

Fig. 1 Skid-steer mobile robot (the UGV) for search and rescue operations developed by BlueBotics (`www.bluebotics.com`) as part of NIFTi project.

2 Theory and methodology

2.1 Extended Kalman Filter

The EKF is a linearized extension of the Kalman filter that estimates the states of a process and their uncertainty from noisy data. The EKF algorithm was implemented using the standard equations (see (Grewal and Andrews 2011, p. 178) for more details and relevant notation in Table 1). Assuming a discrete-time system model disturbed by normally distributed noise:

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

$$\mathbf{z}_k = h_k(\mathbf{x}_k) + \mathbf{v}_k, \qquad \mathbf{v}_k \sim N(0, \mathbf{R}_k) \quad (2)$$

Table 1 Nomenclature

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

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

$$\mathbf{x}_0^+ = E[\mathbf{x}_0^+], \qquad \mathbf{P}_0^+ = E[(\mathbf{x}_0 - \mathbf{x}_0^+)(\mathbf{x}_0 - \mathbf{x}_0^+)^T] \tag{3}$$

The EKF algorithm can be described in two steps using the following discrete-time equations: 1) Time update: Predicted (a priori) state is computed; state transition Jacobian is determined to evaluate the predicted covariance:

$$\mathbf{x}_k^- = f_{k-1}(\mathbf{x}_{k-1}^+, \mathbf{u}_{k-1}), \quad \mathbf{F}_{k-1} \approx \frac{\partial f_k}{\partial \mathbf{x}} \mid_{\mathbf{x}_{k-1}^+, \mathbf{u}_{k-1}} \tag{4}$$

$$\mathbf{P}_k^- = \mathbf{F}_{k-1}\mathbf{P}_{k-1}^+\mathbf{F}_{k-1}^T + \mathbf{Q}_{k-1} \tag{5}$$

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

$$\mathbf{H}_k \approx \frac{\partial h_k}{\partial \mathbf{x}} \mid_{\mathbf{x}_k^-}, \quad \mathbf{K}_k = \mathbf{P}_k^-\mathbf{H}_k^T(\mathbf{H}_k\mathbf{P}_k^-\mathbf{H}_k^T + \mathbf{R}_k)^{-1} \tag{6}$$

$$\mathbf{x}_k^+ = \mathbf{x}_k^- + \mathbf{K}_k(\mathbf{z}_k - h_k(\mathbf{x}_k^-)), \ \mathbf{P}_k^+ = (\mathbf{I} - \mathbf{K}_k\mathbf{H}_k)\mathbf{P}_k)^- \tag{7}$$

$$\mathbf{x}_k^+ = \mathbf{x}_k^- + \mathbf{K}_k(\mathbf{z}_k - h_k(\mathbf{x}_k^-)), \ \mathbf{P}_k^+ = (\mathbf{I} - \mathbf{K}_k\mathbf{H}_k)\mathbf{P}_k)^- \tag{8}$$

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

2.2 Process Models

Four different schemes for data fusion using EKF are presented in this paper and shown in Fig. 2 (see relevant notation in Table 1). Two different system process modeling approaches were investigated the nonlinear model (NLM) and the error model (ERM), see Fig.2 a and Fig. 2 c for implementation without the CF (standard solutions), and Fig. 2 b and Fig. 2 d for the implementation with the CF for attitude estimation (denoted as the grey block in Fig. 2). The CF was introduced in (Kubelka and Reinstein 2012) and proved to be an effective algorithm, where the attitude determination is based on fusion of gravity vector determined from specific forces via a coarse alignment algorithm (see (Titterton and Weston 1997, p. 282)), and Euler angles (roll, pitch and yaw) computed as integration of angular rates. First, the nonlinear model (for details see (Nemra and Aouf 2010)) is based on coordinate frames transformation (where the nonlinearities are caused by the direction cosine matrix) and numerical integration of the IMU outputs. It handles navigation states (position in navigation frame, velocity in body-frame and Euler angles) directly. For the purpose of UGV navigation, the centripetal acceleration corrections can be neglected.

Second, the error model, derived by linearizing the nonlinear differential equations using perturbation analysis and thoroughly described in (Shin 2005), is based on a 15-state concept, expanding the estimates of position and velocity in navigation frame, and attitude errors (i.e., the difference from the expected value) with the biases of the inertial sensors (the actual sensor errors). The data fusion must be completed with an error control loop, which provides corrections to the navigation states computed using the differential navigation

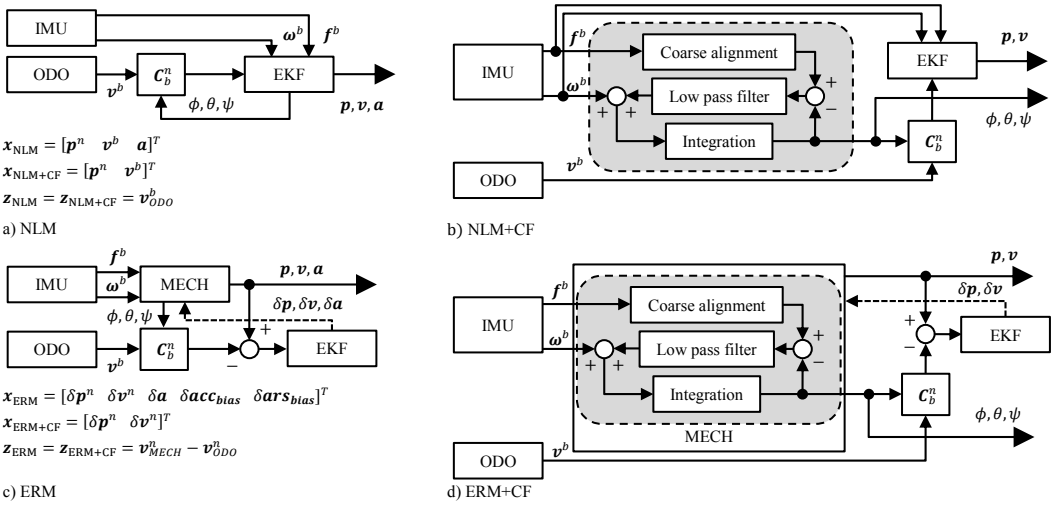

Fig. 2 IMU and odometry (ODO) data fusion schemes for the EKF: a) with nonlinear model (NLM), b) with nonlinear model and CF for attitude estimation (NLM+CF), c) with mechanization (MECH) and error model (ERM), d) with mechanization (MECH), error model and CF for attitude estimation (ERM+CF); x and z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

equations named as mechanization by (Rogers 2003, p. 74), Unlike in the case of the nonlinear model, it is necessary to implement the mechanization, because the error model provides only corrections, not the actual navigation states. Every time the corrections are fed back to the mechanization, the EKF state vector must be set to zeroes. Considering the solutions enhanced by the proposed CF (Kubelka and Reinstein 2012), which provides stable estimates in pitch and roll channels, there is no need to estimate the attitude angles (or attitude corrections) and biases of the gyroscopes within the EKF model, since the feedback of the complementary filter assures stability and drift suppression. Therefore, in both the nonlinear and the error model with the CF, the state vector consists of position and velocity estimates or their respective error estimates.

2.3 Measurement Model

The measurement model is based on the standard body frame odometry model transformed to the navigation frame (see Table 1 for details about frames). Since the UGV is bounded to the surface, it is said to be governed by nonholonomic constraints, which can be exploited to aid the estimation of IMU alignment (Dissanayake et al 2001). These constraints are incorporated into the measurement model, such that no side slip and no vertical movement (such as free fall) is assumed along the body frame lateral y-axis and vertical z-axis, respectively. Therefore, 3D odometry model is implemented as follows: the tracks and UGVs body are aligned, therefore the navigation frame velocity is computed as

$$\mathbf{v}_k^n = \mathbf{C}_b^n \mathbf{v}_k^b = \mathbf{C}_b^n [\,0.5(\mathbf{v}_L + \mathbf{v}_R)\;\; 0\;\; 0\,]^T \tag{9}$$

where $\mathbf{v}_L, \mathbf{v}_R$ are the velocities of the left and right tracks respectively, and the nonholonomic constraints are incorporated as zero values in the y and z component of the displacement vector. The body to navigation frame transformation matrix is defined as:

$$\mathbf{C}_b^n = \begin{bmatrix} c_\theta c_\psi & s_\phi s_\theta c_\psi - c_\phi s_\psi & c_\phi s_\theta c_\psi + s_\phi s_\psi \\ c_\theta s_\psi & s_\phi s_\theta s_\psi + c_\phi c_\psi & c_\phi s_\theta s_\psi - s_\phi c_\psi \\ -s_\theta & s_\phi c_\theta & c_\phi c_\theta \end{bmatrix} \tag{10}$$

where ϕ, θ, ψ are the roll, pitch and yaw angles (Euler angles), and $c_\theta = \cos\theta$, $s_\theta = \sin\theta$ etc. The EKF measurement vector for both the NLM and the NLM+CF, see Fig. 2 a, consists of velocity in the navigation frame. In case of the ERM and the ERM+CF, the measurement vector is produced as the difference between the velocity computed by the mechanization, and odometry-determined velocity, both expressed in navigation frame, as shown in Fig. 2 c. However, the wheel slip along the body frame x-axis is still apparent, and the model cannot compensate for it.

2.4 Observability Analysis

Observability analysis determines, whether a state can be estimated from available measurements (Farrell 2008,

p. 85). Observability rank tests were implemented according to (Weiss et al 2012; Shin 2005) along with the inspection of behavior of the estimation error covariance matrix. The observability analysis of the 9 navigation states in both the NLM and the ERM revealed the following conclusions about the attitude (also corresponds to results in (Weiss et al 2012)): roll and pitch angles are observable with respect to navigation frame due to gravity measurement provided by IMU; the yaw angle becomes unobservable with respect to the navigation frame when the system becomes stationary. Based on these conclusions, an odometry derived motion constraint was introduced to reduce the drift in the unobservable yaw. As the odometry directly indicates when the UGV is stationary, an artificial limit to constant heading was utilized and proved to eliminate the drift in the unobservable yaw angle. Since the velocity is provided to the EKF as measurement, its covariance converges quickly; therefore the velocity estimation error is bounded. This does not reflect the cases, when the UGV is exposed to maneuvers on harsh terrain causing slippage; however, compensating it is not aim of this work (we address it separately in (Reinstein et al 2013)). Consequently, the position related covariance diverges in all four approaches, as it should in any other dead reckoning approach that lacks (in principle) the absolute position measurements (Aghili and Salerno 2013). This is covered in detail in (Grewal and Andrews 2011, p. 277) and confirms that additional position measurements are necessary to achieve observable position estimates.

2.5 Filter Initialization, Tuning, and Consistency

Initialization sequence requires the UGV to be stationary for some time. During this interval, the coarse alignment algorithm (see (Titterton and Weston 1997, p. 282)) is used to determine roll and pitch angles from averaged specific forces. Along with the initial attitude, biases of the gyroscopes are estimated as the average of the static angular rate measurements. This requires at least 1 minute averaging and optimally 3 minutes (providing appropriate bias estimates for 20 minute experiments). Initial position of the UGV always corresponds with origin of the local navigation frame and initial velocity is set to zero, as well as their initial covariance matrices. Filter consistency is defined as convergence of the estimate to the true value (BarShalom et al 2001, p. 232). If the estimator is inconsistent, estimates are not reliable, because their accuracy is unknown. As the process of choosing noise covariance matrices (referred as filter tuning) is often based on the tradeoff between obtaining consistent filter and small estimation errors (BarShalom et al 2001, p. 243), the noise covariance matrix was tuned such that both real-time consistency tests involving the zero-mean innovation test and the innovation whiteness test (BarShalom et al 2001, p. 240) were passed. Subsequently, the process noise covariance matrix was adjusted to achieve small estimation error with respect to the ground truth. Consistency of the position estimate was checked via the inspection of the covariance and the progress of position error (see an example for a 2D experiment using the NLM+CF approach in Fig. 3). The summary of the EKF initialization parameters (state vector, error covariance matrix, process noise covariance matrix and measurement noise covariance matrix) is given in Table 5.

3 Experimental Evaluation

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

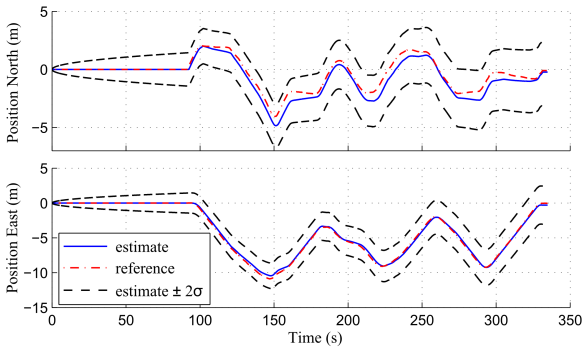

Fig. 3 Position and standard deviation estimates used for filter tuning with respect to the position reference; one of the 2D tracked experiments as obtained using the NLM+CF approach.

The experiments involved four types of field-testing environments namely the 2D tracked outdoor experiments (planar; evaluated with respect to ground truth

Table 2 Experiment datasets

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

trajectories[3] available only for the 2D tracked dataset), indoor experiments on hallway (UGV exploring aground floor corridor inside a building) and stairway (extended to climbing of wooden staircase and descending at slope of 20 deg), and outside environment (long-term driving through a natural environment including hillside and stair climbing). All datasets are referred to using the corresponding labels. In the case of 3D experiments, ground truth was not available and hence the UGV was driven back to the exact starting position each time. Therefore, the return position error and the return attitude error were evaluated instead of RMSE with respect to a reference. The total distance travelled during the 2D experiments was 2.2 km (35 experiments), and 2 km during the 3D experiments (5 experiments for each 3D field-test). Average characteristics of each experiment type are concluded in Table 2.

3.1 A. Example of Results: The Stairway Experiment

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

Attitude estimates that were obtained during the stair descent using all the approaches are shown in Fig. 4 (please note, that the NLM+CF and the ERM+CF exploit the same CF for attitude estimation, thus the figure shows only the CF attitude instead). There are minor differences in roll and pitch channels for the NLM and the CF approaches. Although the ERM performs locally inconsistently, overall stability is ensured and depends mainly on the filter tuning. The

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

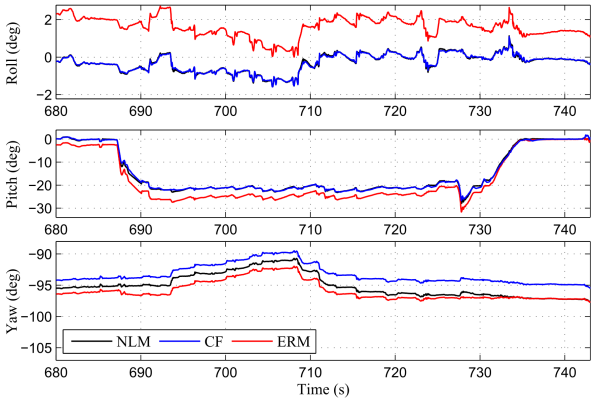

Fig. 4 Euler angles for the stairway experiment stair descent (680 to 745 seconds zoom) as obtained using all attitude approaches

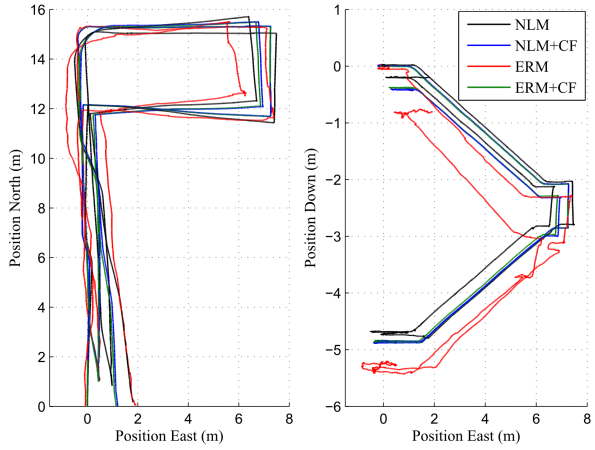

Fig. 5 North-East (x-y) and Down-East (z-y) projections of the 3D trajectory for the stairway experiment as obtained using all four approaches

major contribution to the position error in dead reckoning rises from the yaw estimates. All approaches provide slightly different yaw angle during the stair descent and overall experiment. The consequence can be seen in the trajectory estimates shown in Fig. 5 (position projected to 2D North-East and Down-East directions), where all the North-East horizontal turns should be nearly perpendicular. In the used dead reckoning, the overall trajectory precision depends mostly on attitude estimates. Another type of positioning error is apparent at the ground and first floor in the Down-East view in Fig. 5, where slipping of the tracks during climbing and descending contributes also to the different vertical position estimates.

Table 3 Average Summary of the Performance Evaluation: 2D RMSE, Return Position Error (RPE) of distance driven, Return Attitude Error (RAE) of roll and pitch Angles

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

Table 4 Computational Load Analysis (MATLAB)

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

Table 5 EKF Initialization Parameters

Nonlinear model	
$\mathbf{x}_0$	$[\mathbf{0}_{6\times1}\ \phi_0\ \theta_0\ \psi_0]^T$
$\mathbf{P}_0$	$\mathtt{diag}([\mathbf{0}_{6\times6}\ 0.05\times\mathbf{I}_{3\times3}])$
$\mathbf{Q}_k$	$\mathtt{diag}(10\times[\sigma^2_{\mathtt{acc}}\mathbf{I}_{3\times3}\ \sigma^2_{\mathtt{gyr}}\mathbf{I}_{3\times3}])$
$\mathbf{R}_k$	$0.1\times\mathbf{I}_{3\times3}$
Error model	
$\mathbf{x}_0$	$[\mathbf{0}_{15\times1}]^T$
$\mathbf{P}_0$	$\mathtt{diag}([\mathbf{0}_{6\times6}\ 0.01\times\mathbf{I}_{9\times9}])$
$\mathbf{Q}_k$	$\mathtt{diag}([\mathbf{0}_{9\times9}\ \sigma^2_{\mathtt{acc}}\mathbf{I}_{3\times3}\ \sigma^2_{\mathtt{gyr}}\mathbf{I}_{3\times3}])$
$\mathbf{R}_k$	$10^{-3}\times\mathbf{I}_{3\times3}$
diag constructs block diagonal matrix; $\sigma^2_{\mathtt{acc}}=5\times10^{-4}\,\mathrm{m}^2/\mathrm{s}^4$ and $\sigma^2_{\mathtt{gyr}}=5\times10^{-4}\,\mathrm{rad}^2$ stand for accelerometer and gyroscope noise variances determined experimentally from static measurements	

3.2 Performance Evaluation

To demonstrate relative performance, all the collected datasets were evaluated statistically, averaging and concluding all the experiments in Table 3. The RMSE in 2D position (determined as the RMSE of the north and east position), the average return position error of distance driven, and the average return attitude error of roll and pitch absolute differences, show that the approaches with the CF outperform the standard ones for both the nonlinear model and the error model implementation.

The computational load analysis of all the four approaches was also performed. The results in Table 4 (evaluated in MATLAB 2012a running on PC with Intel Core i7 2.6GHz, 8GB RAM) indicate that the most demanding error model with mechanization consumed approximately 2 times more computational time than the nonlinear model. In the NLM+CF and the ERM+CF, the state vectors are reduced as discussed in Section 2.2, thus the EKF computational load is lower, especially for the ERM+CF. According to the results concluded in Table 3 and Table 4, the NLM+CF approach was selected as the best candidate for the onboard processing in ROS Fuerte version WillowGarage (2012b).

4 Implementation Details

Since the reader should benefit not only from the knowledge gained from presented experimental evaluation, this section provides more general remarks about the implementation. The NLM can be considered as the most straightforward EKF navigation algorithm, based on suboptimal equations estimating position in the navigation frame, velocity in the bodyframe and attitude in Euler angles. Thus, this process model ensures straightforward debugging and EKF tuning. The ERM and mechanization represents computationally demanding but precise navigation algorithm designed to cope with coning and sculling effects in inertial sensors, Coriolis and gravity corrections, Earths rate and transport rate (please consult terms with (Shin 2005) and (Titterton and Weston 1997, p. 335)). The reader should consider the mechanization with ERM when desiring such complex solution providing position and velocity corrections in the navigation frame and attitude corrections as a direction cosine matrix compensation. On the other hand, the filter tuning has to be performed experimentally and is directly connected with the precision of the ground truth and number of experiments. Using inertial sensors only, the CF proved to provide stable and reliable roll and pitch angles for the UGV platform, where no significant centripetal acceleration occurs (Kubelka and Reinstein 2012). The combination with the above mentioned NLM and ERM allows to exploit attitude estimation with the EKF maintaining reduced state vector consisting only of position and velocity estimates or corrections, respectively. ROS is a cross-platform middle-ware that uses publish-subscribe messaging system for data exchange and communication over TCP sockets between processes (ROS nodes), regardless the machine they are running on. Therefore, the real-time implementation consisted of a single C++ node subscribing to wheel encoders and IMU readings (published asynchronously over a CAN bus), computing the NLM+CF algorithm and publishing the navigation data for further utilization. The internal ROS time-stamps were used for proper process synchronization and an OpenCVMatx template class for appropriate matrix management and operations (WillowGarage

2012a). The average EKF cycle duration of final real-time implementation of the NLM+CF approach onboard the UGV (Intel Core2Quad CPU Q9100 2.26 GHz, 8GB RAM) was approximately 40 μs.

5 Conclusion

This paper presents a comparison and experimental evaluation of four different EKF-based estimation architectures for dead reckoning of a mobile robot providing data from inertial measurement unit and wheel encoders. The dead reckoning was realized using the EKF with the nonlinear model and the error model, both with and without a complementary filter for attitude estimation. Field testing in indoor and outdoor environment was carried out to select the best candidate in terms of navigation performance and computational load. According to the results, the nonlinear model with complementary filter was selected as the best approach, running 2 times faster than the error modelbased approaches, and reaching 0.8 m RMSE and average of 4 % return position error of distance driven. It was then implemented in C++ for onboard processing in ROS. With this algorithm implemented, the robot was then deployed in a real USAR mission in Mirandola, Italy (for further details see (Kruijff et al 2012)).

References

Aghili F, Salerno A (2013) Driftless 3-d attitude determination and positioning of mobile robots by integration of imu with two rtk gpss. IEEE/ASME Transactions on Mechatronics 18(1):21–31

Arsenault A, Velinsky SA, Lasky TA (2011) A low-cost sensor array and test platform for automated roadside mowing. IEEE/ASME Transactions on Mechatronics 16(3):592–597

BarShalom Y, Li XR, Kirubarajan T (2001) Estimation with Applications To Tracking and Navigation. Wiley and Sons

Cook G (2011) Mobile robots: Navigation, Control and Remote Sensing. John Wiley & Sons, Inc.

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, DOI 10.1109/70.964672

Endo D, Okada Y, Nagatani K, Yoshida K (2007) Path following control for tracked vehicles based on slip-compensating odometry. In: Proc. IEEE/RSJ Int. Conf. Intelligent Robots and Systems IROS 2007, pp 2871–2876, DOI 10.1109/IROS.2007.4399228

Farrell JA (2008) Aided Navigation: GPS with High Rate Sensors. McGraw-Hill

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, DOI 10.1109/TVT.2011.2105896

Grewal MS, Andrews AP (2011) Kalman filtering: theory and practice using MATLAB. John Wiley & Sons

Kruijff GJM, Pirri F, Gianni M, Papadakis P, Pizzoli M, Sinha A, Pianese E, Corrao S, Priori F, Febrini S, Angeletti S, Tretyakov V, Linder T (2012) Rescue robots at earthquake-hit mirandola, italy: a field report. In: Proceedings of the 10th IEEE International Symposium of Safety Security and Rescue Robotics

Kruijff, G-J et al (2014) Experience in system design for human-robot teaming in urban search & rescue. In: Proceedings of 8th International Conference on Field and Service Robotics, Spring Verlag, STAR, pp 111–125

Kubelka V, Reinstein M (2012) Complementary filtering approach to orientation estimation using inertial sensors only. In: Proc. IEEE Int Robotics and Automation (ICRA) Conf, pp 599–605

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: Proc. IEEE Int Robotics and Automation (ICRA) Conf, pp 619–626

Mobedi B, Nejat G (2012) 3-d active sensing in time-critical urban search and rescue missions. IEEE/ASME Transactions on Mechatronics 17(6):1111–1119

Morales Y, Tsubouchi T, Yuta S (2009) Vehicle 3d localization in mountainous woodland environments. In: Proc. IEEE/RSJ Int. Conf. Intelligent Robots and Systems IROS 2009, pp 3588–3594

Nemra A, Aouf N (2010) Robust ins/gps sensor fusion for uav localization using sdre nonlinear filtering. IEEE Sensors Journal 10(4):789–798, DOI 10.1109/JSEN.2009.2034730

Reinstein M, Hoffmann M (2011) Dead reckoning in a dynamic quadruped robot: Inertial navigation system aided by a legged odometer. In: Proc. IEEE Int Robotics and Automation (ICRA) Conf, pp 617–624, DOI 10.1109/ICRA.2011.5979609

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

Rogers RM (2003) Applied Mathematics in Integrated Navigation Systems. American Institute of Aeronautics and Astronautics, Inc.

Shen J, Tick D, Gans N (2011) Localization through fusion of discrete and continuous epipolar geometry with wheel and imu odometry. In: Proc. American Control Conf. (ACC), pp 1292–1298

Shin EH (2005) Estimation techniques for low-cost inertial navigation. PhD thesis, University of Calgary, Calgary

Song X, Seneviratne LD, Althoefer K (2011) A kalman filter-integrated optical flow method for velocity sensing of mobile robots. IEEE/ASME Transactions on Mechatronics 16(3):551–563

Titterton DH, Weston JL (1997) Strapdown Intertial Navigation Technology. Lavenham: The Lavenham Press ltd

Vasconcelos JF, Cardeira B, Silvestre C, Oliveira P, Batista P (2011) Discrete-time complementary filters for attitude and position estimation: Design, analysis and experimental validation. IEEE Transactions on Control Systems Technology 19(1):181–198

Weiss S, Achtelik M, Chli M, Siegwart R (2012) Versatile distributed pose estimation and sensor self-calibration for an autonomous mav. In: Robotics and Automation (ICRA), 2012 IEEE International Conference on, pp 31 –38, DOI 10.1109/ICRA.2012.6225002

WillowGarage (2012a) Opencv. URL http://opencv.org/

WillowGarage (2012b) Ros. URL http://ros.org/
Yi J, Zhang J, Song D, Jayasuriya S (2007) Imu-based localization and slip estimation for skid-steered mobile robots. In: Proc. IEEE/RSJ Int. Conf. Intelligent Robots and Systems IROS 2007, pp 2845–2850

Chapter 5

Improving Multi-modal Data Fusion by Anomaly Detection

5.1 Summary of the contributions

This chapter covers mainly the second and third contributions of the thesis (see Chapter 3). We build on the previous work, where the inertial and odometry measurements are combined with the exteroceptive sources of complementary information for incremental pose estimation—a camera and laser scanner [Kubelka et al., 2015]. Such a rich sensory set is prone to different nature of errors—exteroceptive modalities are influenced or even corrupted by the dynamics of the vehicle and surrounding environment.

As the fusion of multiple sensory data generally leads to more robust and precise solution, the individual exteroceptive modalities have to be investigated for their correctness. The common practice related to EKF monitoring is to apply standard statistical tests on the observations or residuals, and reject those measurements that do not pass the tests. In practice, detecting such anomalies is a complicated task and therefore, here lies our major contribution. We go beyond simple thresholds or statistical tests by exploring different machine learning approaches and by using our rich datasets[1] with ground truth data. We also show that this extension handles serious local disturbances and rejection of these disturbances vastly improves the overall performance, even in case of operating in challenging environment. In this work we focus on following objectives:

[1]The datasets are publicly available at https://sites.google.com/site/kubelvla/public-datasets

- multi-modal data fusion extension to exteroceptive measurement monitoring (anomaly detection)
- comparison of the common practice approaches and different supervised machine learning approaches
- demonstration of the anomaly detection effectiveness in challenging indoor and outdoor environment

5.2 Publication

The work is represented by a publication with original formatting and follows on the next page.

Autonomous Robots
(The final publication is available at Springer via http://dx.doi.org/10.1007/s10514-015-9431-6)

Improving multi-modal data fusion by anomaly detection

Jakub Simanek · Vladimir Kubelka · Michal Reinstein

Received: 25 June 2014 / Accepted: 5 January 2015

Abstract If we aim for autonomous navigation of a mobile robot, it is crucial and essential to have proper state estimation of its position and orientation. We already designed a multi-modal data fusion algorithm that combines visual, laser-based, inertial, and odometric modalities in order to achieve robust solution to a general localization problem in challenging Urban Search and Rescue environment. Since different sensory modalities are prone to different nature of errors, and their reliability varies vastly as the environment changes dynamically, we investigated further means of improving the localization. The common practice related to the EKF-based solutions such as ours is a standard statistical test of the observations—or of its corresponding filter residuals—performed to reject anomalous data that deteriorate the filter performance. In this paper we show how important it is to treat well visual and laser anomalous residuals, especially in multi-modal data fusion systems where the frequency of incoming observations varies significantly across the modalities. In practice, the most complicated part is to correctly identify the actual anomalies, which are to be rejected, and therefore here lies our major contribution. We go beyond the standard statistical tests by exploring different state-of-the-art machine learning approaches and exploiting our rich dataset that we share with the robotics community. We demonstrate the implications of our research both indoor (with precise reference from a Vicon system) as well as in challenging outdoor environment. In the final, we prove that monitoring the health of the observations in Kalman filtering is something, that is often overlooked, however, it definitively should not be.

Keywords localization · Kalman filter · multi-modal data fusion · anomaly detection · mobile robots

J. Simanek
Department of Measurement, Faculty of Electrical Engineering, Czech Technical University in Prague, Technicka 2, 16627 Prague, Czech Republic

V. Kubelka
Department of Cybernetics, Faculty of Electrical Engineering, Czech Technical University in Prague, Karlovo namesti 13, 12135 Prague, Czech Republic

M. Reinstein
Department of Cybernetics, Faculty of Electrical Engineering, Czech Technical University in Prague, Karlovo namesti 13, 12135 Prague, Czech Republic
E-mail: reinstein.michal@fel.cvut.cz

1 Introduction

In this work, we seek for an improvement in localization of a mobile robot (see Fig. 1) by inspecting measurements provided to a state estimation algorithm. We wish to achieve more accurate and robust estimates of the robot pose (i.e., attitude, velocity, and position). We seek for a classifier that identifies an anomalous measurement and thus supports the pose estimation, which is in our case realized using an extended Kalman filter (EKF) combining proprioceptive (inertial measurements and velocities of the tracks) and exteroceptive (visual and laser measurements) modalities. In the context of state estimation, we define the anomalous measurements as those, which cause significant deviation of the state estimates from the true values. Consequently, anomaly detection is understood as a process that ensures identification and rejection of these defective exteroceptive pose data before they are processed by the estimation framework and spoil the localization. Since we use both statistical and machine learning methods,

the key to successful evaluation of our anomaly detection is having rich datasets with precise and accurate external reference for attitude and position, such as captured using the Vicon system (for this purpose we exploit our publicly available datasets, see Section 4.1).

To avoid confusion, we would like to stress out that our multi-modal data fusion utilizes only relative observations (raw pose measurements or increments). Neither proprioceptive nor exteroceptive pose measurements exploit any kind of predefined map, loop closures or other prior knowledge that could correct the pose information before it is processed in the EKF update step. We primarily aim to improve the localization by mitigating the pose drift that accompanies any such data fusion without absolute corrections. Moreover, we consider the proprioceptive sensors—inertial measurement unit (IMU) and track odometry—as the primary sensory set for the mobile robot localization, which was previously proved to localize the robot up to a certain performance limit influenced especially by the yaw angle drift and track slippage (Reinstein et al 2013; Simanek et al 2014). To complement the proprioceptive set, we exploit the exteroceptive measurements, which provide corrections leading to more robust estimates. We claim that inspecting the Kalman filter residuals (i.e., difference between predicted and real measurements) generated by the exteroceptive sensory set and rejecting real-world disturbances is particularly essential for long-term navigation and autonomous behavior no matter which platform is used.

During our previous work in urban search and rescue (USAR) scenarios, we encountered frequent abnormal patterns in the visual odometry attitude (VO), and laser range finder attitude and position estimates. These usually occurred as a consequence of unexpected environmental effects or modality failures (e.g., dynamically changing conditions, terrain obstacles, limited range of view, low visibility). For this reason, we were motivated to investigate means of detection and subsequent rejection of such anomalous measurements and we extend our previous work regarding EKF-based mobile robot localization and adaptive odometry (Kubelka et al 2014; Simanek et al 2014; Reinstein et al 2013; Reinstein and Hoffmann 2013; Kruijff et al 2012). Furthermore, the aspect of inspecting EKF residuals is often overlooked in common practice or a simple threshold based approach is applied. In that case, threshold has to either be manually tuned for all of the EKF states or implemented as a statistical measure (e.g., chi-squared statistics test). In both cases, the detection performance was repeatedly referred as arguable (Ting et al 2007; Ma et al 2012).

Our major contribution lies in the in-depth analysis of the detection and resolution of anomalous measurements in EKF-based multi-modal data fusion framework. We propose several possible solutions using supervised machine learning: Gaussian Mixture Models (GMM) modified by a Mahalanobis distance-based decision, one-class Support Vector Machines (SVM), and logistic regression. We compare them to each other as well as to the common practice approaches—covariance and chi-squared threshold tests. We evaluate classifier performance by the means of true positive rate and true negative rate on the indoor datasets labeled with the use of a precise and accurate referential system. Finally we deploy the best classifier in several complex testing environments, such as urban park, disaster training site or dense forest, and evaluate the localization performance.

This paper is structured as follows. Related work is addressed in Section 2. Section 3 briefly introduces the data fusion used in our mobile robot and summarizes methods we use for anomaly detection. Section 4 provides details regarding experimental setup used for gathering data, as well as results achieved. The results are concluded in Section 5.

Fig. 1 The skid-steer mobile robot. See Section 3 for details about the sensory set and multi-modal data fusion.

2 Related work

In this overview we focus on two areas: mobile robot localization—multi-modal data fusion without any prior knowledge about the surroundings or absolute measurements (e.g., GPS measurements, known map or landmarks, loop closures); and anomaly detection within the context of state estimation.

2.1 Multi-modal localization

Since our mobile robot is a ground vehicle, the most used and basic means of localization is the wheel or track 2D odometry. Odometry usually provides 2D pose estimates without excessive computational load at the cost of unbounded and significant error growth (Endo et al 2007). Since the standalone odometry drifts rapidly (especially in the case of a skid-steered vehicle) and provides only planar localization, it is often complemented by inertial measurements (Yi et al 2007), slippage estimation (Endo et al 2007), or odometry derived constraints (Dissanayake et al 2001). Generally, it is desired to aid the proprioceptive pose estimates by exteroceptive measurements to provide more accurate and reliable results, but still with unbounded error growth.

Visual odometry (Scaramuzza and Fraundorfer 2011) acts as one of the exteroceptive modalities and provides aiding for the IMU and odometry fusion (Shen et al 2011); however, at the high computational costs that have to be investigated (Rodriguez F et al 2009). Another type of such sensor modality is the laser scanner, which can provide dense 3D point clouds used for mapping, or scan-to-scan translation and rotation estimates allowing robot pose estimation (Pomerleau et al 2013; Yoshida et al 2010; Suzuki et al 2010).

The well established and popular (Görner and Stelzer 2013; Yi et al 2007; Dissanayake et al 2001) state estimation algorithm—the Kalman filter—excels in working with a uni-modal distribution parametrized by mean and covariance, providing state as well as its uncertainty estimates (BarShalom et al 2001, pp. 381–394). Majority of the mobile robotics application use the extended or unscented Kalman filters (UKF), where the UKF compared to EKF does not require the derivation of Jacobians. The computational complexity of the UKF is higher (Chen 2003), which makes it less feasible for running the state estimation in real-time and on-board for large state space. Particle filters are a suitable alternative for localization especially for non-linear systems with non-Gaussian noise. However, they require even higher computational power than the UKF (Gustafsson et al 2002).

2.2 Anomaly detection in the context of state estimation

Anomaly detection addresses the task of finding patterns that do not conform to expected behavior (Chandola et al 2009). There has been a lot of effort put into rejection of such disturbances in a broad range of different applications (e.g., network security, fraud alert, medical domain, etc. Dua and Du 2011).

Within the field of robotics the problem of system failures identification is often referred to as *fault detection and isolation* (FDI) (Pettersson 2005). Fault is defined as an anomaly in behavior of the monitored system and can be detected (indication that something is going as not expected), isolated (locating the faulty component) and identified (determination of magnitude of the fault) (Gertler 1998, p. 3). A reliable robotic system must deal with many uncertainties that can be handled by FDI, e.g., by quality inspection of the provided information (Brunner et al 2013), comparing information providers (Sundvall and Jensfelt 2006), using information flow between the control and actuation (Christensen et al 2008), monitoring the reliability of resources (Morales et al 2008), or proper recognition and modeling of the sensor and mechanical failures (Goel et al 2000).

Failures in the exteroceptive perception systems are one of many sources of uncertainty in mobile robot localization. In visual odometry or laser point cloud processing, outliers often occur in the frame-to-frame or scan-to-scan motion estimation process. It is crucial to implement outlier removal before the motion estimation step (Howard 2008) and monitor the image (point cloud) quality to prevent perceptual failures, which cause large localization errors (Brunner et al 2013). An established standard for model estimation in the presence of outliers is represented by the RANSAC based algorithms (Fraundorfer and Scaramuzza 2012; Konolige et al 2011). Despite the search for outliers is not the concern of our paper, a brief review has to be given to distinguish them from our exteroceptive-based pose anomalies we detect.

The most common fault detection approaches in robotics employ observer-based monitoring of the system and one of the most popular observers is represented by the Kalman filter (Pettersson 2005). Remedies for faulty measurements provided to the Kalman filter may include adjustment of the filter noise matrices (Sarkka and Nummenmaa 2009; Borges and Aldon 2003), smoothing algorithms (Agamennoni et al 2011) or other modifications. These aim to make the Kalman filter estimates more robust by considering heavy-tailed non-Gaussian distributions, re-sampling or weighting techniques (Ting et al 2007; Borges and Aldon 2003). Most of these modifications are still dependent on the undesired data detection, noise identification, or require modified versions of the Kalman filtering algorithms. Traditional statistical methods used for Kalman filter monitoring are based on the filter innovations (i.e., residual between the real and estimated measurement) (Hwang et al 2010; Soule et al 2005) and assume that the residuals are zero-mean Gaussian processes with

given covariance. When the filter is optimal and consistent (BarShalom et al 2001, p. 232), residuals can be checked if they are generated from a chi-squared distribution (Ali and Ushaq 2009; Caron et al 2006; Sukkarieh et al 1999). However, in practice one may struggle to tune the filter properly, since not all the filter assumptions are met, as reported for instance during fusion of VO and IMU attitude (Ma et al 2012).

There exist machine learning alternatives to the standard inspection of the measurements. These are applied in many areas (Chandola et al 2009), such as trajectory tracking and monitoring (Laxhammar et al 2009), or human activity recognition monitored by classifier ensembles (Sagha et al 2013). The idea of combining anomaly detection and state estimation appeared also in the area of filter self-tuning for automated CPU usage monitoring (Knorn and Leith 2008).

It was pointed out by Ndong and Salamatian (2011) that anomaly detection under the assumption about Kalman filter innovations (i.e., white and Gaussian process) produce large amount of false alarms and that innovations should be rather considered a mixture of Gaussians. In general, data driven anomaly detection exploits both supervised (utilizing data labeled as normal and anomalous) and unsupervised methods of learning (especially clustering and nearest neighbor methods) (Tsai et al 2009).

3 Methodology

In the following methodology sections we, first, describe measurements provided by the robot sensory set. Second, we review the data fusion framework and provide a brief summary of the EKF equations and filter tuning (see Kubelka et al (2014) for further explanation). Third, we present the anomaly detection concept along with both statistical and machine learning methods.

3.1 Proprioceptive measurements

Proprioceptive measurements are in our case provided by inertial measurement unit (IMU) and track odometry.

The IMU uses gyroscopes and accelerometers to provide angular rates and accelerations at 90 Hz in all three dimensions. Processing of the IMU data includes integration of angular rates to obtain attitude information and double integration of accelerations to obtain velocity and position, respectively. Both are known to be exposed to a time related drift, however gyroscope integration holds over longer periods of time and can be aided by accelerometers in the roll and pitch channels to suppress the drift (Kubelka and Reinstein 2012). In contrast, accelerometer integration is always prone to severe drift, therefore usage of the position estimates is restricted to very short time intervals and we primarily correct this drift by incorporating track odometry.

Second type of proprioceptive sensor, incremental rotary encoder, provides us the velocity of right and left tracks at 15 Hz. Since the odometry heading information degrades quickly due to skid-steer character of driving, we incorporate only the speed in the robot forward direction in our model. The slippage still affects both velocity and position, and if not corrected, these quantities are prone to unbounded error growth.

The main disadvantage of combining IMU and odometry lies in unbounded yaw and spatial drifts. Since we do not use any absolute pose measurements, we can only rely on other relative sources of aiding. Because the exteroceptive measurements are usually exposed to different sources of errors, we exploit them to aid the attitude and position estimates.

3.2 Exteroceptive measurements

Exteroceptive measurements are provided by panoramic camera and rotating laser scanner.

Camera images are processed by visual odometry (VO) (Scaramuzza and Fraundorfer 2011; Svoboda et al 1998; Divis 2013), which provides us relative rotation between two consecutive panoramic images sampled at 2.5 Hz. Implementation uses `OpenCV`[1] Orb keypoint detector, 5-point RANSAC and sliding window bundle adjustment refinement (Kümmerle et al 2011). In the case of omnidirectional camera, the most needed yaw angle aiding is well conditioned, since the image correspondences are significant in the majority of the omnidirectional image. However, there are many occasions where the VO attitude deteriorates: changing environment, camera occlusion, loss of image correspondences due to low-textured surfaces, insufficient or extensive illumination of the scene, or inadequate VO bandwidth with respect to robot dynamics (e.g., quick turns).

High density 3D point clouds produced by the rotating laser scanner are processed by the Iterative Closest Point (ICP) algorithm (Besl and McKay 1992; Pomerleau et al 2013) (implementation uses `libpointmatcher`[2]). It provides us relative translation and rotation between two consecutive scans sampled at 0.3 Hz. As in the case of VO, there are situations, where

[1] Open-source computer vision library `http://opencv.org/`.

[2] Open-source ICP library `https://github.com/ethz-asl/libpointmatcher`.

the laser information may fail or deteriorate, especially in harsh environment, where an obstacle may cause the rotating laser to stop, or frequent indoor/outdoor transitions or large moving objects occur. The main problem is the low sampling rate of the laser pose information, which is in our case the only modality that can correct the error accumulated during the track slippage.

3.3 Data fusion scheme

In our multi-modal data fusion (originally published in Kubelka et al (2014)), we exploit the EKF in feedback form (Farrell 2008, p. 209) to fuse several sensor modalities sampled at significantly different frequencies to estimate robot pose in a local coordinate frame (see Fig. 2, the modalities are highlighted in gray).

The robot is modeled as a rigid body moving through space with no dissipative forces. Following sensors are used in our data fusion: IMU Xsens MTi-G containing low-grade 3-axis accelerometer and 3-axis gyroscope; caterpillar track odometer; omni-directional Point Grey Ladybug3 camera; and continuously rotating SICK LMS-151 laser range finder capable of creating 3D scans of the environment in front of the robot.

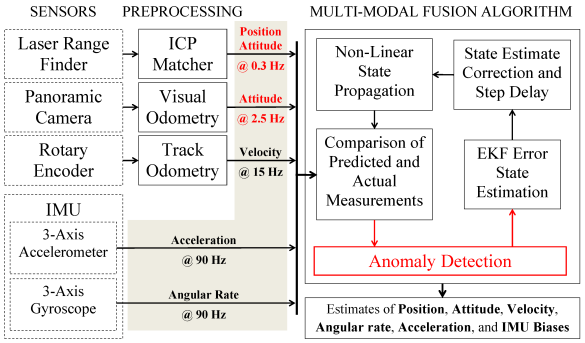

Fig. 2 EKF based multi-modal data fusion algorithm (Kubelka et al 2014) with integrated anomaly detection (anomaly detection, VO and laser modalities marked in red)

3.3.1 Extended Kalman filter

The EKF in feedback form follows the error state implementation inspired by Weiss (2012) where the error state $\Delta\mathbf{x}$ is defined as the difference between true state and state estimate $\Delta\mathbf{x} = \mathbf{x} - \hat{\mathbf{x}}$. A flowchart summary of the EKF equations is given in Fig. 3 and can be split into following steps:

- Nonlinear State Propagation: The *a priori* state estimate $\hat{\mathbf{x}}_{k|k-1}$ is obtained using a discretized nonlinear system model f and previous state estimate $\hat{\mathbf{x}}_{k-1|k-1}$.
- Error Prediction Step: The error state and error state covariance propagation in time are computed as follows: the *a priori* error state estimate $\Delta\mathbf{x}_{k|k-1}$ is initialized to zeroes and *a priori* error state covariance $\mathbf{P}_{k|k-1}$ is estimated using previous error state covariance, system matrix $\mathbf{F}_d$ and system noise covariance matrix $\mathbf{Q}_d$ (both discretized by the Van Loan method (Farrell 2008, p. 143)).
- Error Update Step: Measurements $\mathbf{y}$ and nonlinear measurement model h (computes predicted measurements based on current *a priori* state estimates) are used to produce the measurement residuals $\mathbf{v}$ (innovations)[3]. Measurement residuals update the *a posteriori* error state estimate $\Delta\mathbf{x}_{k|k}$ via Kalman gain matrix $\mathbf{K}$. Kalman gain is computed using the residual matrix $\mathbf{S}$, measurement noise covariance matrix $\mathbf{R}$, measurement matrix $\mathbf{H}$, and error state covariance matrix $\mathbf{P}$. Final update step involves updating the error state covariance matrix to the *a posteriori* form $\mathbf{P}_{k|k}$.
- State Estimate Correction: The EKF cycle is completed by correcting the *a priori* state with the updated error state to obtain *a posteriori* state $\hat{\mathbf{x}}_{k|k}$.

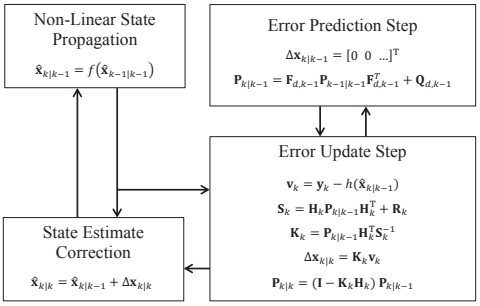

Fig. 3 Error state EKF flowchart (Kubelka et al 2014).

3.3.2 Filter tuning

The filter tuning (i.e., choosing the system and measurement noise covariances) was carried out in two phases ensuring proper filter performance. First, noise covariances were estimated according to the measured noise under static conditions or noise levels specified in the manufacturer's technical specification. Second, the

[3] We exploit this difference between predicted and actual measurements in the anomaly detection.

noise covariances were iteratively tuned until the filter provided the best possible state estimates with respect to the precise and accurate reference obtained by Vicon motion capture system (see Section 4.1).

We are aware that the consistency criteria (zero-mean white innovations with given covariance (Simon 2006, pp. 298–301)) may not be passed and the estimated covariance may end up too pessimistic or optimistic. However, such use of the referential data in the tuning process is considered a common practice and is an alternative approach to time-consuming manual approach (Lau and Lin 2011). It also ensures the best possible localization performance in the given environment.

3.4 Anomaly detection

3.4.1 Detection and rejection of anomalous exteroceptive data

It is highly advised to inspect the sensor data and reject the detected anomalous readings in the Kalman filter framework (Grewal and Andrews 2008, p. 271). The detection of anomalous data was in our case done by both statistical tests and machine learning approaches (see Section 3.4.2). All classifiers were deployed for the real-time detection in the EKF framework and when the anomaly was detected, the system ignored the measurement that caused the anomalous behavior.

To detect anomalies by statistical tests in real-time, we employed the information provided solely by the EKF. First, we tracked the error state estimates and looked for values larger than their estimated uncertainty. Second, we applied a statistical measure called chi-squared test, which monitors the innovation information provided by the EKF.

The machine learning approaches using supervised offline learning require the ground truth labels defining whether the exteroceptive measurements (i.e., VO attitude, laser attitude and position) are normal or anomalous. First, we ran the EKF framework with every experiment available in our datasets (see Section 4.1 for more details about the datasets). Second, referential pose from the Vicon motion capture system was compared with the EKF state estimates after the particular measurement was used for the EKF update. In other words, we artificially matched the estimate to the true value and tracked the progress of the precision of estimates. Finally, if the state estimate started to diverge rapidly from the true value in a short time interval (varying according to the modality sampling time), the measurement was labeled as anomalous; the other measurements were labeled as normal. Thus we ended up with the normal data within a specific tolerance around the true value during the fixed time interval.

Choosing the right input for the classifiers (i.e., features) is key in supervised learning. As our anomaly detection approach is purely based on EKF residuals, we propose features that consist of residuals in each dimension for particular quantity of interest (VO attitude, lased attitude and position) as well as their norm. Where whole residuals capture more the spatial dependency of the difference between actual and predicted measurements, the norm clearly indicates the distance of measurements from their prediction. For each classifier we performed a straightforward feature selection from the following set: residuals, norm of residuals, both norm and residuals. Best performing subset of the features was selected for the classification task (see Table 2).

Last but not least, there is a necessary prerequisite of having sufficiently large datasets for learning the hypothesis (i.e., the model that ensures detection of the anomaly). In our case, we divided the experiments (see Section 4.1 for details) as follows: 2/3 of the experiments were used as a training set (for learning) and 1/3 was used as a testing set (this part of the dataset served also for evaluation of the statistical methods).

3.4.2 Classification methods

This section introduces the methods we implemented for anomaly detection.

State estimates and variance are checked by monitoring diagonal terms of the covariance matrix $\mathbf{P}$, which correspond to estimation uncertainty of the particular error states (Farrell 2008, pp. 217–224). Since the covariance matrix, unlike the state estimates, is not dependent on the actual measurement data, the covariance analysis can be used for anomaly detection after measurement update of the filter. If the error state estimate is larger than a certain threshold (typically the $\pm 3\sigma$ boundary), the measurement is classified as anomalous.

Chi-squared gating test (referred also as a *filter consistency check*) is a standard procedure used in combination with the Kalman filter. It utilizes either the normalized estimation error squared (NEES), which however requires knowledge about the true states, or the normalized innovation squared (NIS) (BarShalom et al 2001, p. 232-244). In practice, there is a need for real-time consistency check, which is often realized as monitoring of the NIS $\epsilon = \mathbf{v}^T \mathbf{S}^{-1} \mathbf{v}$, where $\mathbf{v}$ is the innovation (residual) vector, and $\mathbf{S}$ is the innovation (residual)

covariance matrix. NIS has a χ^2 distribution with l degrees of freedom (i.e., dimension of $\mathbf{v}$) (BarShalom et al 2001, p. 236) and the test is performed by comparing the NIS to a value from the χ^2 distribution table with given confidence level and degree of freedom.

Gaussian mixture models (GMM) is a probabilistic approach that models the data as a mixture of different Gaussian distributions. There are k Gaussian distributions parametrized by their centers μ, covariance matrices $\mathbf{\Sigma}$, and mixture proportions. The solution of finding the mixture of distributions with k components was implemented as maximum-likelihood parameter estimation and found by the Expectation-Maximization algorithm (Mitchell 1997, pp. 191–195).

In our approach, we estimated the GMM (i.e., learned their parameters) on the training data divided into anomalous and normal samples. Both were modeled by two separate mixtures and each mixture was modeled with a number of up to three components. For the classification, we computed the Mahalanobis distance of a incoming residual to every Gaussian in both anomalous and normal GMM. The Mahalanobis distance is defined as $d_M = \sqrt{(\mathbf{v}-\mu)^T\mathbf{\Sigma}^{-1}(\mathbf{v}-\mu)}$. The distances were then weighted by the mixture proportions of the GMM components (i.e., their estimated probability), averaged to obtain weighted distance measures (indicating the distance of new residual to trained anomalous and normal GMM), and the final decision was made by comparing these two distance measures. [4]

One-class SVM algorithm was introduced by Schölkopf et al (1999) as a support vector algorithm for *novelty detection*. The one-class SVM assumes that the available data are from one class, thus it can learn only on the normal data (majority of the training dataset) and distinguish the anomalies without knowing them. The problem is specified in following way: given a dataset with a probability distribution P, estimate a subset S of the feature space, such that a previously unseen point drawn from P lies outside the subset S with some a priori bounded probability. One-class SVM solves this problem by estimating a function that is positive on S (captures most of the normal data) and negative on the complement (anomalous data). The one-class version of SVM is characterized by the parameter ν which is an upper bound on the fraction of outliers and a lower bound on the fraction of support vectors. As the one-class SVM is sensitive to the chosen parameters (Sagha et al 2013), we also compared the default settings of the selected kernels [5] with the parameters found by a grid search: fraction parameter ν (i.e., $\nu = 2^{-10}, 2^{-9.5}, \ldots, 2^{-1}$; for all approaches); RBF parameter ($\gamma = 2^{-10}, 2^{-9}, \ldots, 2^{0}$) and degree of the polynomial kernel ($d = 0, 1, \ldots, 6$). The grid search used a 5-fold cross-validation (according to SVM guide for practice (Hsu et al 2003)) and we included the anomalous samples in the validation data to obtain the training error (evaluated in the same way as defined in Section 4.1) on a dataset including both classes. [6]

Logistic regression can solve a binary classification problem (Murphy 2012, pp. 245–249), where model parameters are estimated by an iterative search for the minimum of the negative log-likelihood. In our case this was solved by the gradient descent algorithm. [7]

4 Experiments and results

Our experimental evaluation consists of two parts. First, we exploit indoor laboratory environment equipped with the Vicon motion capture system to gather datasets with precise ground truth[8] for the classifier learning and performance evaluation. Second, we show the impact of the best anomaly detection method on the accuracy of actual localization in four selected test cases from both indoor and outdoor unstructured environment.

4.1 Classification performance

For learning of the classifiers and evaluation of their classification performance, we tried to simulate USAR environment, including ramps, boxes, catwalks, small passages, etc. (see Fig. 4 that shows part of the setup). We recorded approximately 1.7 km of indoor data with ground truth [9]; 20 runs represent standard conditions (590 m in total), 25 runs represent failure cases of visual and laser modalities (e.g., partially blocked field of

[4] We used the *gmdistribution class* from Matlab Statistics toolbox and fitted the GMM for number of components equal to $k = 1, 2, 3$ with the standard parameters. Maximum of 3 mixtures was selected, because models with more than 3 mixtures usually resulted in negligible weights for some redundant mixtures.

[5] We took linear, polynomial and radial basis function (RBF) kernels in consideration.

[6] We used the LIBSVM tool (version 3.17) (Chang and Lin 2011).

[7] We used the *fminunc* from Matlab Optimization toolbox for the minimization.

[8] Data collected in a room monitored with twelve cameras covering more than 20 m^2 and giving a few millimeter accuracy at 100 Hz

[9] The datasets are publicly available at `https://sites.google.com/site/kubelvla/public-datasets`

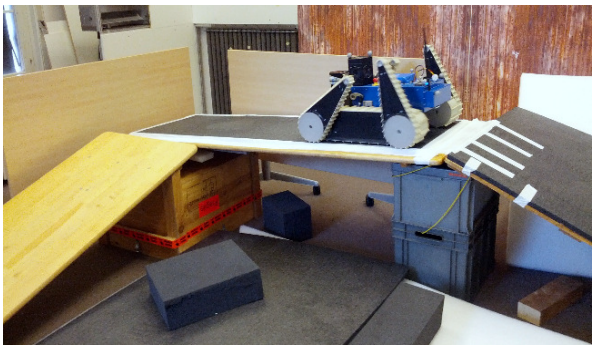

Fig. 4 The environment simulating USAR conditions (slippery slopes, catwalks etc.)

Table 1 Learning datasets. Number of all anomalous vectors, all residual vectors, and anomaly ratio.

Modality	Anomalies	Residuals	Anomaly ratio (%)
VO attitude	354	16871	2.1
Laser attitude	91	3477	2.6
Laser position	23	3477	0.7

view, moving object in the field of view; 1120 m in total). The number of all learning data for VO and laser modalities is shown in Table 1, where the anomaly ratio is the ratio of anomalies to the normal data.

4.1.1 Classification performance metrics

As our approach is a binary classification problem, we evaluated the performance of all classifiers by means of a standard two-by-two confusion matrix[10]. The commonly used classification metrics (i.e., accuracy, error rate) are however not suitable for evaluating our imbalanced dataset (see the anomaly ratio in Table 1). We aimed for maximizing true positive rate $\mathrm{TPR} = \mathrm{TP}/(\mathrm{TP}+\mathrm{FN})$, as well as true negative rate $\mathrm{TNR} = \mathrm{TN}/(\mathrm{TN}+\mathrm{FP})$. We used these two classification metrics to maximize the percentage of positive and negative examples correctly classified at the same time in the form of a geometric mean measure $\mathrm{G} = \sqrt{\mathrm{TPR} \cdot \mathrm{TNR}}$ (Kubat et al 1997).

4.1.2 Classification results

Table 2 shows performance of all classifiers over the following modalities: VO attitude, laser attitude and position. The results were evaluated on the testing dataset, which was randomly selected from the whole dataset, thus the results show generalization of the methods to new data.

Performance of the statistical methods (covariance analysis and chi-squared test) exhibit relatively good results in the VO attitude modality, however with a high number of FP. This behavior is not such an issue, since the roll and pitch angles are primarily determined using the IMU measurements at high rate. The high number of FP causes covariance analysis to have the lowest TPR of all classifiers in the attitude modalities. Thus, it tends to incorrectly reject the normal data—the thresholds, defined implicitly by the EKF tuning of process and measurement noise covariances, seem to be set too low. However, as empirically verified on all the testing data, this behavior cannot be fixed by simply increasing the threshold, because then the TNR is improved at the expense of TPR. The lower TPR in the chi-squared test indicates that this classifier can miss real anomalies and let them propagate. This can be harmful especially in the yaw angle estimation, where no absolute corrections are available. In general, statistical methods perform poorly in the case of laser modalities and even fail in the laser position modality. The main reason behind it is the nature of the laser position measurements, which aim to correct the large error accumulated in track odometry due to slipping. When the laser measurement is processed, large values are introduced in the position residuals, thus any threshold-based method will classify such correction as anomalous data.

Two of the machine learning methods—GMM (learned with 2 mixtures) and one-class SVM (OCSVM, learned with RBF kernel and parameters determined via grid search)—outperform the other classifiers in all modalities tested. On top of that, the number of FP was lower in case of the GMM in the laser modalities. Low number of FP is in general more significant for the laser than visual modality, mainly due to the low sampling frequency of the laser measurements.

Logistic regression classifier does not achieve such performance, because there is only about 2 % of anomalies present in the datasets and regression should have reasonably balanced classes for proper learning. However, we wanted to include these results to show, how the straightforward and popular learning approach will perform.

4.1.3 Discussion on classification performance

We can conclude that both the covariance analysis and the chi-squared threshold test were inferior to machine learning techniques we proposed and both are implicitly

[10] True Positive (TP) – anomaly correctly classified as anomaly; False Negative (FN) – anomaly incorrectly classified as normal; False Positive (FP) – normal data incorrectly classified as anomaly; True Negative (TN) – normal data correctly classified as normal.

Table 2 Classification results as obtained on the testing examples

	Features	TP	TN	FP	FN	TPR	TNR	G
VO attitude								
COV	$\Delta\mathbf{x}_{\mathtt{att}}$, $\sigma_{\mathtt{att}}$	157	6773	971	12	0.93	0.87	0.901
CHI	$\mathbf{v}_{\mathtt{att}}, \mathbf{S}_{\mathtt{att}}$	130	6867	877	39	0.77	0.89	0.826
GMM	$\|\mathbf{v}_{\mathtt{att}}\|$	160	7702	42	9	0.95	0.99	**0.970**
SVM	$\mathbf{v}_{\mathtt{att}}$	160	7701	43	9	0.95	0.99	0.970
REG	$\|\mathbf{v}_{\mathtt{att}}\|$	47	7743	1	122	0.27	0.99	0.527
Laser attitude								
COV	$\Delta\mathbf{x}_{\mathtt{att}}$, $\sigma_{\mathtt{att}}$	26	1170	407	1	0.96	0.74	0.845
CHI	$\mathbf{v}_{\mathtt{att}}, \mathbf{S}_{\mathtt{att}}$	14	1429	148	13	0.52	0.91	0.685
GMM	$\|\mathbf{v}_{\mathtt{att}}\|$	25	1499	78	2	0.93	0.95	0.938
SVM	$\mathbf{v}_{\mathtt{att}}$	26	1473	104	1	0.96	0.93	**0.948**
REG	$\|\mathbf{v}_{\mathtt{att}}\|$	4	1574	3	23	0.15	0.99	0.385
Laser position								
COV	$\Delta\mathbf{x}_{\mathtt{pos}}$, $\sigma_{\mathtt{pos}}$	0	1596	0	8	0.00	1.00	0.000
CHI	$\mathbf{v}_{\mathtt{pos}}$, $\mathbf{S}_{\mathtt{pos}}$	0	1596	0	8	0.00	1.00	0.000
GMM	$\|\mathbf{v}_{\mathtt{pos}}\|$	8	1574	22	0	1.00	0.99	**0.993**
SVM	$\mathbf{v}_{\mathtt{pos}}$	8	1488	108	0	1.00	0.93	0.965
REG	$\|\mathbf{v}_{\mathtt{pos}}\|$	1	1595	1	7	0.12	0.99	0.353

Note: $\Delta\mathbf{x}$ is the EKF error state vector, $\mathbf{v}$ is the EKF residual vector, $\mathbf{S}$ is the EKF residual covariance matrix, σ is the EKF standard deviation; $\|\mathbf{x}\|$ denotes Euclidean norm of a vector $\mathbf{x}$; subscripts `att` and `pos` denote attitude and position elements in a vector or matrix.
COV – Covariance analysis; CHI – Chi-squared test; GMM – Gaussian Mixture Models with 2 mixtures (best of 1, 2 and 3 mixtures); SVM – One-class SVM with RBF kernel (best of linear, polynomial and RBF kernels; learned using grid search); REG – Logistic Regression.

sensitive to the filter tuning. However, they were not so demanding when searching for the best model. The only parameter tuned was the sigma bound in case of covariance analysis. Also, there is actually a drawback when implementing the covariance test—it requires *a posteriori* error state estimate and hence the EKF measurement update needs to be recomputed when anomaly is detected. Since we deal with a multi-modal data fusion, the pose is usually influenced by other modalities and therefore we cannot simply skip the measurement step.

Both the GMM and the SVM methods performed in a very similar way and generalized for the indoor dataset (the performance on training and testing datasets was very similar). However, the number and range of parameters that needs to be tuned is larger in case of the SVM (i.e., bounding parameter ν and different kernels with associated parameters) than in case of the GMM, where we consider only the number of Gaussian mixtures. Thus the implementation of the GMM was more straightforward and computationally less demanding.

Finally, we would like to comment on the importance of fault isolation and identification—the two stages subsequent to detection. In our case, it is not necessary and often even not possible to identify the exact cause of the anomalous exteroceptive measurements since the measurements are already processed by the VO or the ICP algorithms. Therefore in general, inspection needs to be carried out at the lower level, where the images or point clouds are processed. Although isolation and identification may help the rescuers and operators to understand the situation, the pose estimates will benefit mainly from the detection part of the FDI process.

Table 3 Summary of the test case experiments

	Test case 1	Test* case 2	Test+ case 3	Test case 4
Dist. traveled (m)	32	327	332	550
Exp. duration (min)	3	17	37	37
Elev. difference (m)	0	12	1	8

* Sum of 4 experiments.
\+ Sum of 3 experiments.

4.2 Localization performance

In this section, we demonstrate the performance in localization when solving the anomaly detection using the GMM (with 2 mixtures modified by a Mahalanobis distance based decision). The GMM was selected as best choice based on the results in Table 2 and advantages mentioned in Section 4.1.3. We evaluate the localization in four different scenarios that are most likely to appear in USAR missions. Three of the test cases take place in outdoor environment and are completely new and unseen in the learning process. Section 4.2.1 introduces the scenario of the robot operating in semi-structured indoor environment. Section 4.2.2 presents four experiments from an outdoor urban environment. Section 4.2.3 presents three experiments from a disaster training site with robot inspecting ruins of a building. Section 4.2.4 demonstrates the performance during an outdoor experiment carried out in a challenging forest environment. Table 3 summarizes the total distance traveled (1241 m in total), experiment duration, and maximum elevation difference for all four test cases.

4.2.1 Test case 1: Indoor with Vicon reference

The selected indoor experiment is a representative from the set of the learning experiments in semi-structured environment (see Fig. 4). We show the results with respect to ground truth measurements in both trajectory and attitude. The robot was exploring the room in an approximately square-shaped path, however the VO experienced failures due to lack of tracked features. As shown in red in Fig. 5, sudden loss and consequent lack

of tracked visual features resulted in wrong VO attitude measurements (represented by roll, pitch, and yaw Euler angles). The data fusion framework with GMM as anomaly detection successfully detected and rejected the anomalous VO measurements (marked as gray vertical lines). This prevented deterioration in the state estimates and ensured improvement in the final position error normalized by the distance traveled (i.e., 6.2 % for data fusion without anomaly detection and 0.8 % for data fusion with anomaly detection). This metric is defined as $e_{final} = ||\mathbf{p}_l - \mathbf{p}_{ref,l}||/d$, where l denotes the last position sample and d is the distance traveled in meters. Accuracy improvement is apparent also in the average position error metric defined as $e_{avg}(k) = \sum_{i=1}^{k} ||\mathbf{p}_i - \mathbf{p}_{ref,i}||/k$, where $1 \leq k \leq$ *total number of samples*. Final e_{avg} for data fusion without anomaly detection is more than 3× larger than for data fusion with anomaly detection, see Fig. 5.

4.2.2 Test case 2: Urban environment

Second test case includes outdoor experiments from an urban environment, namely one street experiment in a urban canyon and three experiments from a city park including dense vegetation[11], see Fig. 6. The ground truth position of the robot in these experiments was tracked by an external reference system Total Station TS15 from Leica Geosystems (further refered to as Leica reference)[12]. Unfortunately, the Leica theodolite does not provide the orientation of the robot and the position ground truth is available only when the robot is in the direct line-of-sight with the theodolite. However, it still ensure precise position measurements even in urban areas, which are often not suitable for reliable GPS ground truth measurements due to limited sky view and multipath effects.

As in the previous case, we show the results from all urban experiments as a projection of the trajectory to the 2D plane accompanied by the average position error plots. The depicted urban experiments are as follows: a rectangular trajectory on the street, bigger and smaller lap in the park including stairs, and straight and slightly leveled path in the park. As shown in Fig. 7, there are improvements in terms of position error in all experiments except one, which did not include any severe anomalies and the position error is comparable with the standard results. In terms of final position error normalized by the distance traveled the results are the following: 5.9 %, 6.1 %, 0.9 % and 17.9 % for standard data fusion, and 0.2 %, 1.9 %, 1.5 % and 0.5 % for data fusion with anomaly detection.

4.2.3 Test case 3: Disaster training site

The third test case demonstrates the anomaly detection during USAR experiments carried out in a scenario where the robot was deployed to a training disaster site of fire and rescue service in Prato, Italy (see Fig. 8 and Fig. 11). The robot was teleoperated to assist firefighters in a training mission, where the goal was to find victims in areas inaccessible or too dangerous for people. Because the ground truth for these experiments was not available, we show particular instances where the applied anomaly detection contributed to the overall localization accuracy and robustness of the estimates. Namely, this test case includes two instances of disaster site exploration and one instance of robot inspecting an underground pipeline.

The state estimation results from the site exploration are shown in Fig. 9 and Fig. 10. The robot was teleoperated along the training site and returned back to initial position. There is a visible improvement in both trajectories when using the anomaly detection. The final position error normalized by the distance driven is for both the training site experiments as follows: 16.0 % and 9.5 % for standard data fusion, and 1.2 % and 2.8 % for data fusion with anomaly detection. The main reason of these improvements is due to rejecting anomalous attitude VO and laser corrections, especially in yaw angle, which otherwise spoil the trajectory. Details of such anomalous measurement rejections are shown in the Euler angles plots in Fig. 9 and Fig. 10.

The inspection experiment took place in a pipeline of diameter slightly larger than the robot, see Fig. 11 (e–f). As there was too dark for the VO modality, it was not used at all. Figure 12 shows the Euler angles during the whole experiment, as well as the Euler angles detail of a moment, when the robot drove out of the tunnel and the laser modality experienced sudden change from very tight area in the pipeline to a large area in front of the tunnel. There is a visible deterioration of the roll and yaw angles around the time sample of 181 s. This particular instance was handled well by the data fusion due to the anomaly detection.

4.2.4 Test case 4: Outdoor forest

The last test case covers by far the most challenging environment from all our datasets. The experiment took place in a hilly forest area (see pictures in Fig. 13). The robot traversed uneven terrain covered

[11] These experiments are publicly available as well at https://sites.google.com/site/kubelvla/public-datasets.

[12] The distance measurements are taken in continuous mode at 7.5 Hz with measurement accuracy about 3 mm.

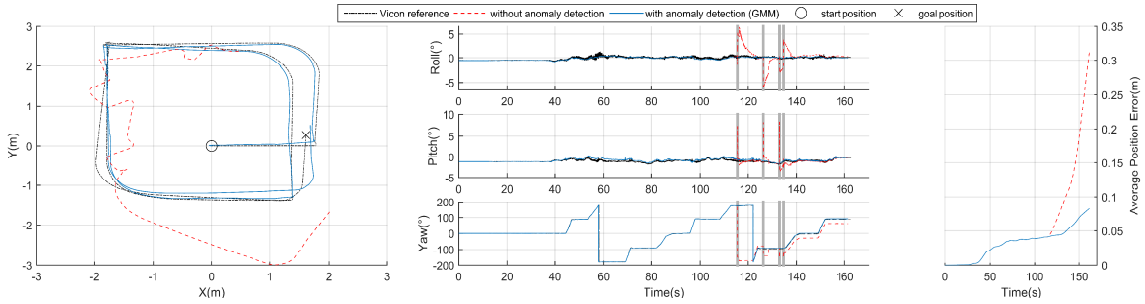

Fig. 5 Indoor experiment - VO attitude failure between 115 s and 140 s (black dash-dot – Vicon reference; red dashed – multi-modal data fusion without anomaly detection; blue – multi-modal data fusion with anomaly detection; gray vertical lines – rejected VO attitude measurements). (Left) Shows 2D projection of the trajectories. (Middle) Euler angles. (Right) Average position error expressed as function of time.

Fig. 6 Pictures from the urban experiments. (Left to Right) Detail of the Leica reference system. Street environment. Park environment. Detail of a staircase in the park.

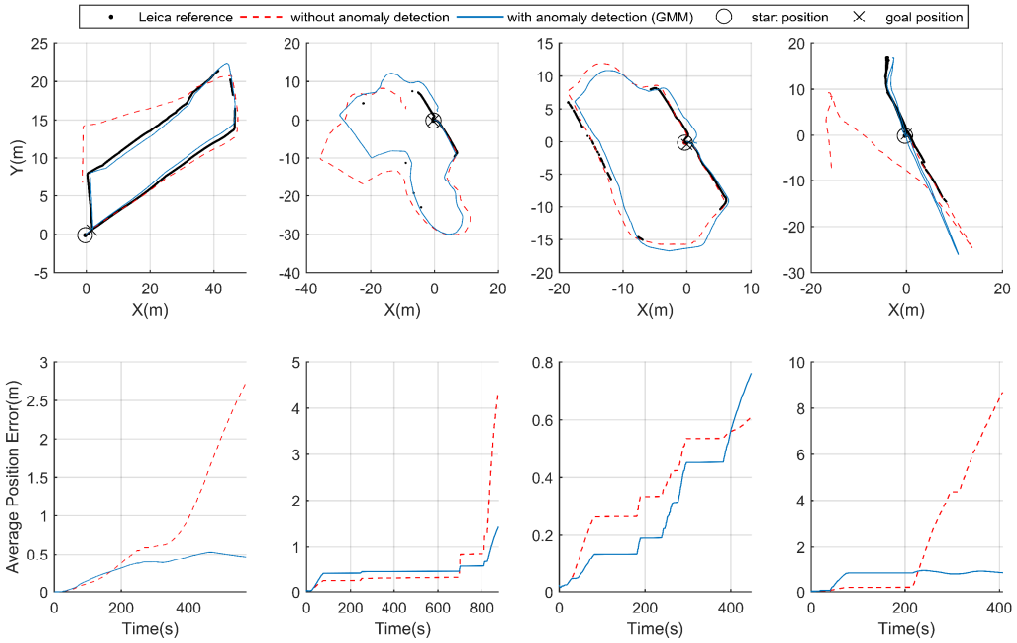

Fig. 7 Urban environment (black dotted – Leica reference; red dashed – multi-modal data fusion without anomaly detection; blue – multi-modal data fusion with anomaly detection). (Top) 2D projections of the trajectories. (Bottom) Average position errors expressed as function of time.

Fig. 8 Pictures from disaster training site experiment.

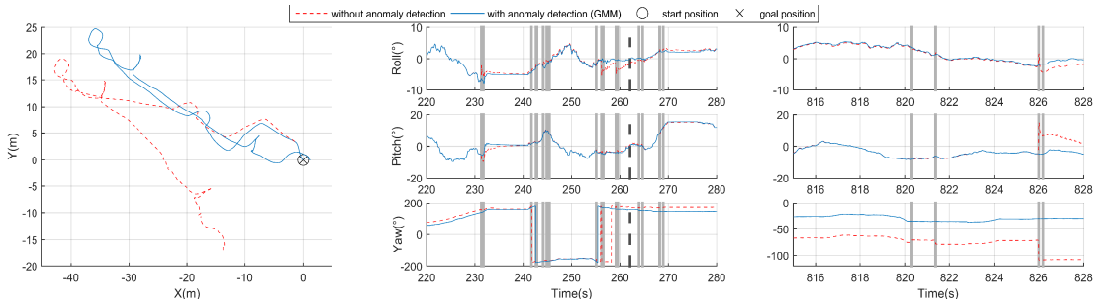

Fig. 9 Exploration of the disaster site 1 (red dashed – multi-modal data fusion without anomaly detection; blue – multi-modal data fusion with anomaly detection; light gray vertical lines – VO anomaly detected and rejected; dark gray dashed vertical lines – laser anomaly detected and rejected). (Left) 2D projection of the trajectories. (Middle) Euler angles detail 1. (Right) Euler angles detail 2.

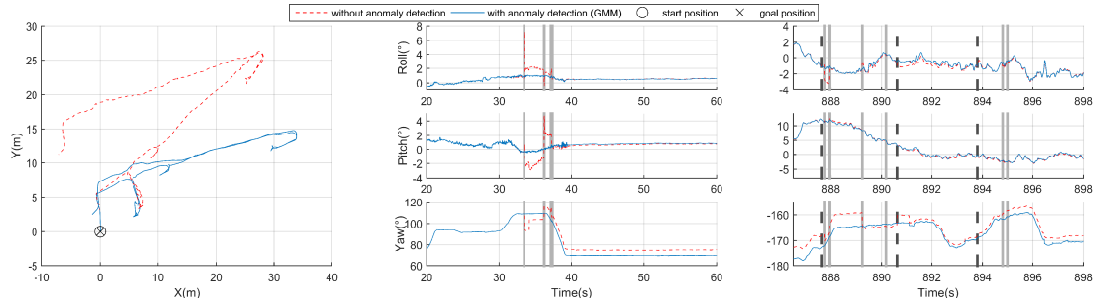

Fig. 10 Exploration of the disaster site 2 (red dashed – multi-modal data fusion without anomaly detection; blue – multi-modal data fusion with anomaly detection; light gray vertical lines – VO anomaly detected and rejected; dark gray dashed vertical lines – laser anomaly detected and rejected). (Left) 2D projection of the trajectories. (Middle) Euler angles detail 1. (Right) Euler angles detail 2.

by soil and foliage causing very high slippage. Moreover, both the vision and laser modalities were exposed to very severe conditions. These were mainly the vegetation blocking both camera and laser range of view, or the robot traversing deformable terrain introducing motions too fast to be properly sampled by relatively low frequency of the exteroceptive modalities. Furthermore, there were people simulating dynamic obstacles moving around the robot and partially blocking the camera and laser scanner range of view during the whole experiment.

As gathering reasonable ground truth data for such experiment is very difficult, we attached a GPS receiver on top of the robot, to get an approximate position information along the robot path. One sigma uncertainty of the GPS position was experimentally determined to be $\pm 5\,\mathrm{m}$ (mainly due to dense vegetation and treetops covering most of the sky). This result indicates that the forest environment is almost GPS-denied case if we aim to achieve localization precision less than the body size of the robot.

A 2D projection of the original trajectory, trajectory obtained with anomaly detection, the referential

Fig. 11 Disaster site experiment from robot point of view. (a) USAR scenario. (b) Entering collapsed basement. (c-d) Inspecting building ruins at the basement level. (e-f) Inspecting a pipeline (scene artificially lightened up).

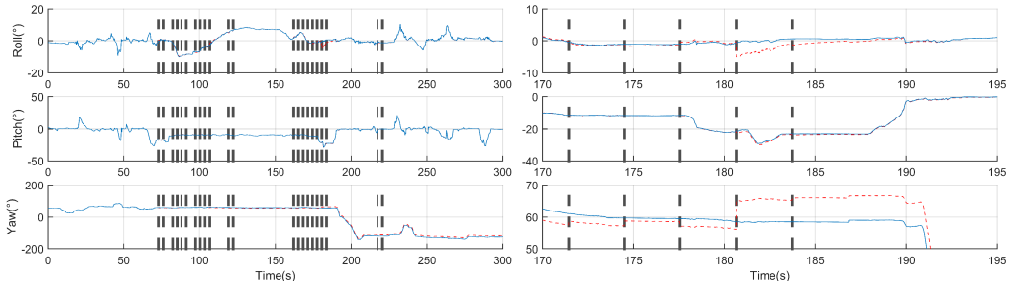

Fig. 12 Inspection of the pipeline (red dashed – multi-modal data fusion without anomaly detection; blue – multi-modal data fusion with anomaly detection; dark gray dashed vertical lines – laser anomaly detected and rejected). (Left) Euler angles.(Right) Detail of the Euler angles around 181s and anomalous laser attitude corrections casued by robot returning from the pipeline and entering large space area.

Fig. 13 Pictures from the forest experiment. (Left) Entering forest on the forest road. (Middle) Climbing a steep hill. (Right) Descending into a valley (model of the robot projected into a point cloud colored using camera)

GPS trajectory with its uncertainty, and the average position errors are shown in Fig. 14. Robot initial position was set to zero coordinates and robot proceeded as indicated by the black arrows. It can be clearly seen by visual inspection that the data fusion with anomaly detection outperformed the standard one. The average position error is computed with respect to the GPS trajectory and the final average error for data fusion without anomaly detection is 2× larger than the error for data fusion with anomaly detection. On the other hand, both trajectories are influenced by a visible yaw estimate drift, which is however expected and common when no absolute yaw corrections are provided.

4.2.5 Discussion on localization performance

We would like to note that in our data fusion approach we do not have any absolute position or heading measurements available—the whole framework is based on relative measurements only—there is no loop-closure in the VO or the ICP mapping algorithm. Therefore, the results presented in Section 4.2 diverge from the reference in the same way as can be expected for a dead reckoning localization approach. Based on the analysis of our experiments, we identified that it is actually the yaw angle that contributes the most to the overall error accumulation in this relative pose estimation. The larger the error in yaw estimates, the larger the position error gets accumulated over time. This effect is

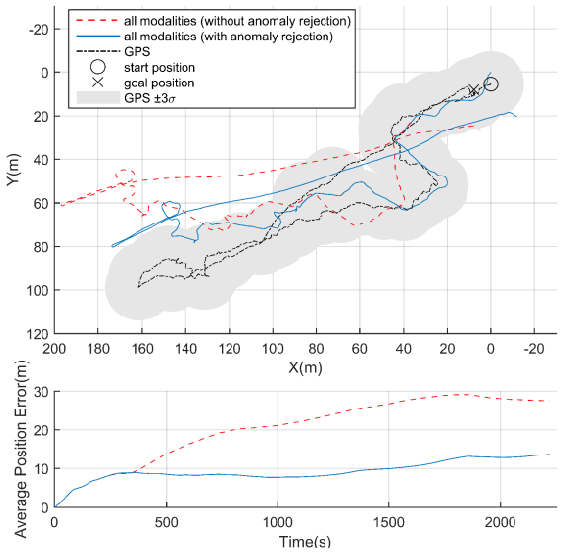

Fig. 14 Forest experiment (black dash-dot – GPS; red dashed – multi-modal data fusion without anomaly detection; blue – multi-modal data fusion with anomaly detection; gray – experimentally determined GPS uncertainty). (Top) 2D projection of the trajectories. (Bottom) Average position error expressed as function of time.

dramatically magnified if anomalies propagate and influence the yaw angle estimates. From the results summarized in Table 4 as well as by visual inspection of all the projected trajectories, we can conclude that the data fusion supported by anomaly detection provides more accurate localization results. The reason is the anomaly detection handles most of the exteroceptive measurements that would otherwise introduce large errors.

In our previous work (Kubelka et al 2014) we introduced a failure case methodology to test the actual limits of our proposed data fusion. Without the anomaly detection, the data fusion algorithm performance was limited as expected. Although it cannot be exactly quantified, we hope that by including the anomaly detection we are able to push these limits and even pass most of the failure cases. Although it cannot be proven, similar behavior can be expected during deployment to similar conditions.

5 Conclusion

In this paper, we extended our previous work regarding the multi-modal data fusion for localization of a mobile robot equipped with inertial sensors, track odometry, omnidirectional camera and laser range-finder. Even if the data fusion was well designed, there still occur

Table 4 Performance summary of the test case experiments with and without anomaly detection (AD). Summary includes all experiments for which the final position error e_{final} or average position error e_{avg} metric is available.

	$e_{final}(\%)$		$e_{avg}(m)$	
	without AD	with AD	without AD	with AD
Indoor	6.2	**0.8**	0.3	**0.1**
Street	5.9	**0.2**	2.7	**0.5**
Park 1	6.1	**1.9**	4.4	**1.4**
Park 2	**0.9**	1.5	**0.6**	0.7
Park 3	17.9	**0.5**	8.7	**0.8**
Train. site 1	16.0	**1.2**	n/a	n/a
Train. site 2	9.5	**2.8**	n/a	n/a
Forest	4.0	**3.4**	27.6	**13.6**

real-world disturbances related to the environment, no matter which platform is used. These disturbances affect significantly the exteroceptive modalities (in our case the vision and the laser) and subsequently deteriorate the state estimates, causing degradation in precision and reliability of localization. We have shown that incorporating the anomaly detection to exteroceptive measurements is an important step towards a more robust and accurate state estimation.

We presented standard statistical tests for monitoring of the EKF observations and compared them to supervised machine learning approaches; namely: the covariance test, chi-squared test, modified Gaussian Mixture Models, one-class Support Vector Machines, and logistic regression. We demonstrated that the GMM modified by a Mahalanobis distance decision criteria performs the best out of the selected classifiers. This was experimentally verified in indoor tests with precise ground truth using a Vicon system, as well as in challenging outdoor environments. The anomaly detection handled serious local disturbances and vastly improved the overall performance, even in case of driving inside a pipe or through a dense forest. Despite the necessity of creating a labeled ground truth—which requires having a precise reference and is a time-costly procedure—for the supervised machine learning, we still think that the improvements are definitely worth the effort put into training the classifier.

Based on the results, we can conclude that anomaly detection is an important stage of any multi-modal state estimation framework and should not be overlooked or underestimated using a simplified approach. We are convinced that by having such reliable localization, we can now pursue our desired quest for long-term autonomy using our framework.

Acknowledgements The research presented here was supported as follows: Jakub Simanek was supported by the Grant Agency of the CTU in Prague (SGS13/144/OHK3/2T/13); Vladimir Kubelka was supported by the Czech Science Foun-

dation (14-13876S); and Michal Reinstein was supported by the European Union FP7 under the TRADR project No. FP7-ICT-609763.

We would like to thank Francis Colas and Francois Pomerleau for their valuable advices concerning the optimal estimation, and Lorenz Oswald for the assistance during the experiments. The Vicon motion capture system was kindly provided by Autonomous Systems Lab, ETH Zurich, Switzerland.

References

Agamennoni G, Nieto JI, Nebot EM (2011) An outlier-robust kalman filter. In: Robotics and Automation (ICRA), 2011 IEEE International Conference on, IEEE, pp 1551–1558

Ali J, Ushaq M (2009) A consistent and robust kalman filter design for in-motion alignment of inertial navigation system. Measurement 42(4):577–582

BarShalom Y, Li XR, Kirubarajan T (2001) Estimation with Applications To Tracking and Navigation. Wiley and Sons

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

Borges GA, Aldon MJ (2003) Robustified estimation algorithms for mobile robot localization based on geometrical environment maps. Robotics and Autonomous Systems 45(3):131–159

Brunner C, Peynot T, Vidal-Calleja T, Underwood J (2013) Selective combination of visual and thermal imaging for resilient localization in adverse conditions: Day and night, smoke and fire. Journal of Field Robotics 30(4):641–666

Caron F, Duflos E, Pomorski D, Vanheeghe P (2006) Gps/imu data fusion using multisensor kalman filtering: introduction of contextual aspects. Information Fusion 7(2):221–230

Chandola V, Banerjee A, Kumar V (2009) Anomaly detection: A survey. ACM Computing Surveys (CSUR) 41(3):15

Chang CC, Lin CJ (2011) LIBSVM: A library for support vector machines. ACM Transactions on Intelligent Systems and Technology 2:1–27, software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm

Chen Z (2003) Bayesian filtering: From kalman filters to particle filters, and beyond. Statistics 182(1):1–69

Christensen AL, OGrady R, Birattari M, Dorigo M (2008) Fault detection in autonomous robots based on fault injection and learning. Autonomous Robots 24(1):49–67

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

Divis J (2013) Visual odometry from omnidirectional camera, master thesis, Charles University in Prague

Dua S, Du X (2011) Data mining and machine learning in cybersecurity. Taylor & Francis

Endo D, Okada Y, Nagatani K, Yoshida K (2007) Path following control for tracked vehicles based on slip-compensating odometry. In: Proc. IEEE/RSJ Int. Conf. Intelligent Robots and Systems IROS 2007, pp 2871–2876

Farrell JA (2008) Aided Navigation: GPS with High Rate Sensors. McGraw-Hill

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

Gertler J (1998) Fault detection and diagnosis in engineering systems. CRC press

Goel P, Dedeoglu G, Roumeliotis SI, Sukhatme G (2000) Fault detection and identification in a mobile robot using multiple model estimation and neural network. In: Robotics and Automation, 2000. Proceedings. ICRA'00. IEEE International Conference on, IEEE, vol 3, pp 2302–2309

Görner M, Stelzer A (2013) A leg proprioception based 6 dof odometry for statically stable walking robots. Autonomous Robots 34(4):311–326

Grewal M, Andrews A (2008) Practical considerations. In: Kalman Filtering:Theory and Practice Using MATLAB, Wiley-IEEE Press, pp 355–426

Gustafsson F, Gunnarsson F, Bergman N, Forssell U, Jansson J, Karlsson R, Nordlund PJ (2002) Particle filters for positioning, navigation, and tracking. Signal Processing, IEEE Transactions on 50(2):425–437

Howard A (2008) Real-time stereo visual odometry for autonomous ground vehicles. In: Intelligent Robots and Systems, 2008. IROS 2008. IEEE/RSJ International Conference on, IEEE, pp 3946–3952

Hsu CW, Chang CC, Lin CJ, et al (2003) A practical guide to support vector classification. Technical report

Hwang I, Kim S, Kim Y, Seah CE (2010) A survey of fault detection, isolation, and reconfiguration methods. Control Systems Technology, IEEE Transactions on 18(3):636–653

Knorn F, Leith DJ (2008) Adaptive kalman filtering for anomaly detection in software appliances. In: INFOCOM Workshops 2008, IEEE, IEEE, pp 1–6

Konolige K, Agrawal M, Sola J (2011) Large-scale visual odometry for rough terrain. In: Kaneko M, Nakamura Y (eds) Robotics Research, Springer Tracts in Advanced Robotics, vol 66, Springer Berlin Heidelberg, pp 201–212

Kruijff GJ, Janicek M, Keshavdas S, Larochelle B, Zender H, Smets N, Mioch T, Neerincx M, Diggelen Jv, Colas F, Liu J, Pomerleau F, Siegwart R, Hlavac V, Svoboda T, Petricek T, Reinstein M, Zimmermann K, Pirri F, Gianni M, Papadakis P, Sinha A, Balmer P, Tomatis N, Worst R, Linder T, Surmann H, Tretyakov V, Corrao S, Pratzler-Wanczura S, Sulk M (2012) Experience in system design for human-robot teaming in urban search & rescue. In: Proceedings of 8th International Conference on Field and Service Robotics, Spring Verlag, STAR, pp 111–125

Kubat M, Holte R, Matwin S (1997) Learning when negative examples abound. In: Machine Learning: ECML-97, Springer, pp 146–153

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

Kubelka V, Oswald L, Pomerlau F, Colas F, Svoboda T, Reinstein M (2014) Robust data fusion of multi-modal sensory information for mobile robots. Journal of Field Robotics In press

Kümmerle R, Grisetti G, Strasdat H, Konolige K, Burgard W (2011) G2o: A general framework for graph optimization. In: Robotics and Automation (ICRA), 2011 IEEE International Conference on, pp 3607–3613

Lau TK, Lin Kw (2011) Evolutionary tuning of sigma-point kalman filters. In: Robotics and Automation (ICRA), 2011 IEEE International Conference on, IEEE, pp 771–776

Laxhammar R, Falkman G, Sviestins E (2009) Anomaly detection in sea traffic-a comparison of the gaussian mixture

model and the kernel density estimator. In: Information Fusion, 2009. FUSION'09. 12th International Conference on, IEEE, pp 756–763

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: Proc. IEEE Int Robotics and Automation (ICRA) Conf, pp 619–626

Mitchell TM (1997) Machine Learning, 1st edn. McGraw-Hill, Inc., New York, NY, USA

Morales Y, Takeuchi E, Tsubouchi T (2008) Vehicle localization in outdoor woodland environments with sensor fault detection. In: Robotics and Automation, 2008. ICRA 2008. IEEE International Conference on, IEEE, pp 449–454

Murphy KP (2012) Machine Learning: A Probabilistic Perspective (Adaptive Computation and Machine Learning series). The MIT Press

Ndong J, Salamatian K (2011) A robust anomaly detection technique using combined statistical methods. In: Communication Networks and Services Research Conference (CNSR), 2011 Ninth Annual, IEEE, pp 101–108

Pettersson O (2005) Execution monitoring in robotics: A survey. Robotics and Autonomous Systems 53(2):73–88

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. Robotics, IEEE Transactions on 29(2):563–571

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

Rodriguez F SA, Fremont V, Bonnifait P (2009) An experiment of a 3d real-time robust visual odometry for intelligent vehicles. In: Proc. 12th Int. IEEE Conf. Intelligent Transportation Systems ITSC '09, pp 1–6

Sagha H, Bayati H, Millan JdR, Chavarriaga R (2013) On-line anomaly detection and resilience in classifier ensembles. Pattern Recognition Letters 34(15):1916–1927

Sarkka S, Nummenmaa A (2009) Recursive noise adaptive kalman filtering by variational bayesian approximations. Automatic Control, IEEE Transactions on 54(3):596–600

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

Schölkopf B, Williamson RC, Smola AJ, Shawe-Taylor J, Platt JC (1999) Support vector method for novelty detection. NIPS 12:582–588

Shen J, Tick D, Gans N (2011) Localization through fusion of discrete and continuous epipolar geometry with wheel and imu odometry. In: Proc. American Control Conf. (ACC), pp 1292–1298

Simanek J, Reinstein M, Kubelka V (2014) Evaluation of the EKF-based estimation architectures for data fusion in mobile robots. Mechatronics, IEEE/ASME Transactions on Early Access

Simon D (2006) Optimal state estimation: Kalman, H infinity, and nonlinear approaches. Wiley. com

Soule A, Salamatian K, Taft N (2005) Combining filtering and statistical methods for anomaly detection. In: Proceedings of the 5th ACM SIGCOMM conference on Internet Measurement, USENIX Association, pp 31–31

Sukkarieh S, Nebot E, Durrant-Whyte H (1999) A high integrity imu/gps navigation loop for autonomous land vehicle applications. Robotics and Automation, IEEE Transactions on 15(3):572–578

Sundvall P, Jensfelt P (2006) Fault detection for mobile robots using redundant positioning systems. In: Robotics and Automation, 2006. ICRA 2006. Proceedings 2006 IEEE International Conference on, IEEE, pp 3781–3786

Suzuki T, Kitamura M, Amano Y, Hashizume T (2010) 6-dof localization for a mobile robot using outdoor 3d voxel maps. In: Proc. IEEE/RSJ Int Intelligent Robots and Systems (IROS) Conf, pp 5737–5743

Svoboda T, Pajdla T, Hlaváč V (1998) Motion estimation using central panoramic cameras. In: Hahn S (ed) IEEE International Conference on Intelligent Vehicles, Causal Productions, pp 335–340

Ting JA, Theodorou E, Schaal S (2007) A kalman filter for robust outlier detection. In: Intelligent Robots and Systems, 2007. IROS 2007. IEEE/RSJ International Conference on, pp 1514–1519

Tsai CF, Hsu YF, Lin CY, Lin WY (2009) Intrusion detection by machine learning: A review. Expert Systems with Applications 36(10):11,994–12,000

Weiss SM (2012) Vision based navigation for micro helicopters. PhD thesis, Diss., Eidgenössische Technische Hochschule ETH Zürich, Nr. 20305, 2012

Yi J, Zhang J, Song D, Jayasuriya S (2007) Imu-based localization and slip estimation for skid-steered mobile robots. In: Proc. IEEE/RSJ Int. Conf. Intelligent Robots and Systems IROS 2007, 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: Proc. IEEE/RSJ Int Intelligent Robots and Systems (IROS) Conf, pp 1414–1420

Chapter 6

Gaussian Processes for Anomaly Detection

In this chapter, we introduce an extension of the work done in Chapter 5—Gaussian Processes (GP) for anomaly detection. First, we present the theory about regression and classification using the GP, describe training of the GP model, and comment on computational complexity and sparse solutions. Then we briefly review related work and introduce the methodology for using GP for the anomaly detection. Lastly, we provide extended and thorough evaluation, where we compare selected GP approaches with each other as well as with methods presented in Chapter 5.

6.1 Introduction

Gaussian processes, known as *kriging*[1] in geostatistics or meteorology [Matheron, 1963], gained a lot of awareness in the machine learning community in recent years, since it is a powerful tool for both regression and classification problems. Essentially, Gaussian processes are a generalization of the Gaussian probability distributions [Rasmussen and Williams, 2006, p. 2]. The GP represent a practical tool to approach the problem of bias-variance trade-off, which can be regarded as the trade-off between model complexity and data-fit, avoiding phenomenons known as *overfitting* and *underfitting* [Briscoe and Feldman, 2011].

We introduce the Gaussian processes in this doctoral thesis, since the nature of the Bayesian framework is different from the methods proposed in Section 5. Mainly, because such a framework allows for advanced decision-making process by using both the predicted output and the uncertainty. This is accompanied by the subjective Bayesian viewpoint that needs to specify knowns and unknowns—place priors over unknown quantities, and infer the posterior distributions over

[1] Process of interpolating and estimating elevation information.

the unknown data [Nowozin, 2014]. The conclusions based on the posterior distributions solve the problem in a slightly different way—the uncertainty of predictions, in terms of predictive GP variance, can actually be used to analyze the regression or classification output. In other words, it tells how close is a new observation to previously observed samples; thus, how reliable are such prediction in the anomaly rejection process.

6.2 Theory and methodology

In this section, we provide a brief introduction to Gaussian Processes in the context of machine learning. We show how the GP can be used in the supervised learning in the form of regression and classification for obtaining real-valued and discrete outputs, respectively. We also provide an additional view into the problem of using Gaussian Processes for anomaly detection. Finally, we present examples of possible learning scenarios and propose different approaches for the anomaly detection applied in an EKF framework.

6.2.1 Gaussian process

Gaussian processes belong to the group of learning algorithms known as the Bayesian methods. Unlike the classical machine learning algorithms, which try to identify the single best model for the training data, the GP compute the posterior distribution over models [Do, 2007]. Such distribution provides additional information in the form of the uncertainty of the predictions, which can help to tell how accurate these predictions are. Since the Gaussian process model is not a parametric model, the effective number of the parameters grows with the training sample size (in contrast with parametric model, which transfers the information from training data into a fixed number of parameters) [Rasmussen, 2004]. Advantage of non-parametric models[2] is that the GP allow to avoid the classical dilemma of choosing a model at appropriate complexity. The Gaussian process model can automatically adapt its complexity to the data, which is an aspect of great practical importance, since it simplifies the training.

A Gaussian Process can be understood as a generalization of a multivariate Gaussian distribution to infinitely many variables. The Gaussian process is a distribution over functions, which is fully specified by a mean function $m(\mathbf{x})$ and covariance function $k(\mathbf{x}, \mathbf{x}')$ of a real process $f(\mathbf{x})$.

[2]Without determining some parameters in advance. Example of such parameters can be found in setting the number of mixtures in Gaussian mixture models.

The mean and covariance functions are defined as

$$\begin{aligned} m(\mathbf{x}) &= \mathbb{E}[f(\mathbf{x})] \\ k(\mathbf{x}, \mathbf{x}') &= \mathbb{E}[(f(\mathbf{x}) - m(\mathbf{x}))(f(\mathbf{x}') - m(\mathbf{x}'))]. \end{aligned} \tag{6.1}$$

where $\mathbb{E}$ denotes the expected value operator.

According to notation from [Rasmussen and Williams, 2006, p. 13], Gaussian process can be written as

$$f(\mathbf{x}) \sim \mathcal{GP}(m(\mathbf{x}), k(\mathbf{x}, \mathbf{x}') \tag{6.2}$$

where $\mathbf{x}$ and $\mathbf{x}'$ are either the training or the test sets, $f = f(\mathbf{x})$ is the latent function value, further concatenated in a vector of latent functions $\mathbf{f} = [f_1, \ldots, f_n]$.

6.2.2 Gaussian process regression

We consider regression as a problem that aims at finding a mapping of input vectors $\mathbf{x} \in \mathcal{R}^D$ to the output $y \in \mathcal{R}$ using labeled training data $\mathcal{D} = \{(\mathbf{x}_i, y_i)|i = 1, \ldots, n\}$. Applying Gaussian Processes for regression is specified by choosing the mean and the covariance function. The regression mapping can be found by Bayesian inference techniques, which are employed within the GP framework [Deisenroth, 2010, p. 9]. Bayesian inference fits the probability model to a set of data and summarizes the result by a probability distribution of the unknown data. According to [Quiñonero-Candela and Rasmussen, 2005], inference in the GP model puts a joint GP prior distribution[3] on the vector of training and test latent values, $\mathbf{f}$ and $\mathbf{f}_*$, and uses Bayes' rule to combine the prior distribution with the likelihood $p(\mathbf{y}|\mathbf{f})$ to obtain the joint posterior distribution

$$p(\mathbf{f}, \mathbf{f}_*|\mathbf{y}) = \frac{p(\mathbf{f}, \mathbf{f}_*)p(\mathbf{y}|\mathbf{f})}{p(\mathbf{y})} \tag{6.3}$$

where $\mathbf{y}$ is the vector of noisy training data. By marginalizing out the unwanted training set of latent variables, the desired posterior distribution is produced as

$$p(\mathbf{f}_*|\mathbf{y}) = \int (p(\mathbf{f}, \mathbf{f}_*|\mathbf{y}))\, \mathrm{d}\mathbf{f} = \frac{1}{p(\mathbf{y})} \int (p(\mathbf{f}, \mathbf{f}_*)p(\mathbf{y}|\mathbf{f}))\, \mathrm{d}\mathbf{f}. \tag{6.4}$$

The whole process can be summarized as follows: first, the prior of the unknown data has to be determined; second, training data are gathered; third, the posterior is computed by updating the prior by incorporating information from the training data.

[3] In this thesis, we abbreviate terms prior and posterior distribution as prior and posterior.

The inference on one dimensional GP prior and posterior for three representations sampled according to method described in [Rasmussen and Williams, 2006, pp. 200-201] is shown in Fig. 6.1. In the this example, left part of the Fig. 6.1 shows the prior samples (representing the

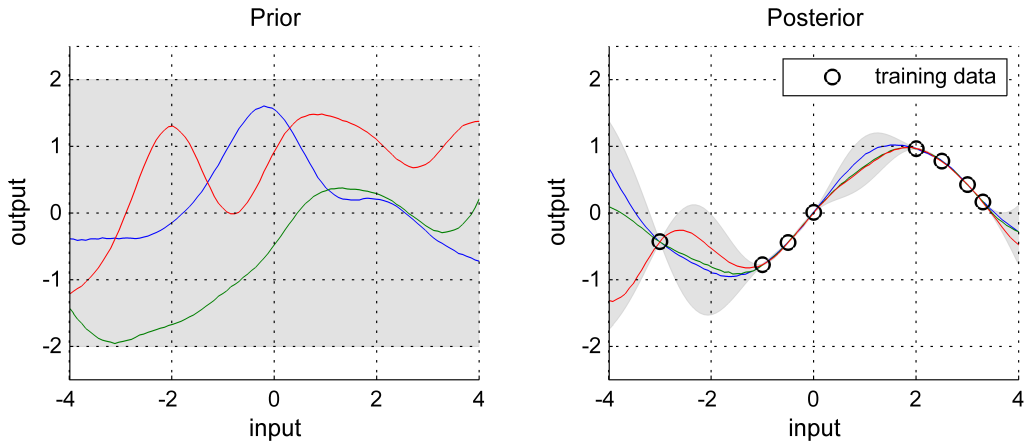

Figure 6.1: Gaussian Process example - prior and posterior. Left plot shows three functions randomly sampled from the GP prior. Right plot shows three functions randomly drawn from the GP posterior. Test data are evenly spaced in the input interval $[-4, 4]$, training data are marked as circles, gray area corresponds to $\pm 2\,\sigma$ interval.

belief without seeing any data) drawn at random from the function f such as the values of f are requested only at finite number of input data. These samples are specified by a GP, which determines specific properties, e.g., it favors smooth functions and has zero mean[4]. It can be clearly seen, that the predicted uncertainty (displayed in gray as $\pm 2\,\sigma$ interval) does not depend in this case on the input data.

To make predictions using the training data, such as in Fig. 6.1 right, the prior has to be updated to obtain posterior. In the simple special case in Fig. 6.1, where the mean function is zero and the measurements are noise free [Rasmussen and Williams, 2006, p. 15], the joint distribution of the training outputs, $\mathbf{f}$, and the test outputs, $\mathbf{f}_*$, according to the prior is

$$\begin{bmatrix} \mathbf{f} \\ \mathbf{f}_* \end{bmatrix} \sim \mathcal{N}\left(\mathbf{0}, \begin{bmatrix} K(X,X) & K(X,X_*) \\ K(X_*,X) & K(X_*,X_*) \end{bmatrix}\right) \tag{6.5}$$

with n training samples and n_* test samples, $K(X, X_*)_{n \times n_*}$ denoting the matrix of covariances for all pairs of training and test samples. By conditioning the joint Gaussian prior distribution

[4]Due to missing knowledge, the mean function is often ignored and set to zero.

on the observation we obtain

$$\begin{aligned}\mathbf{f}_*|X_*, X, \mathbf{f} \sim \mathcal{N}(&K(X_*, X)K(X, X)^{-1}\mathbf{f}, \\ &K(X_*, X_*) - K(X_*, X)K(X, X)^{-1}K(X, X_*))\end{aligned} \tag{6.6}$$

Three representations of the function values $\mathbf{f}_*$ (see Fig. 6.1 right), corresponding to test inputs $\mathbf{X}_*$, are sampled from this joint posterior distribution by evaluating the mean and covariance matrix from Eq. (6.6). The uncertainty in the posterior is reduced around the training samples and falls back to the prior as the test samples go further from training points. This can be clearly seen from Eq. (6.6), where the second positive term is subtracted from the first term, which represents the prior in the covariance, telling how much the training data corrected the belief. Predictions for unseen test samples can be then computed from this posterior.

Many application, in contrast with the example in Fig. 6.1, do not have noise free training inputs. We adopt the most common assumption of additive, independent, identically distributed Gaussian noise ϵ with variance σ_n^2, which describes the noise training data $y = f(\mathbf{x}) + \epsilon$ and is easily taken into account in the GP. The prior on the noisy training data is summed with the noise term $\sigma_n^2 I$.

Following the definition of a GP in Eq. (6.2), the corresponding posterior process including the noise term determines the predictive equations for Gaussian process regression defined as

$$\begin{aligned}\mathbf{f}_*|X, \mathbf{y}, X_* &\sim \mathcal{GP}(\bar{\mathbf{f}}_*, \mathrm{cov}(\mathbf{f}_*)) \\ \bar{\mathbf{f}}_* &= K(X_*, X)[K(X, X) + \sigma_n^2 I]^{-1}\mathbf{y} \\ \mathrm{cov}(\mathbf{f}_*) &= K(X_*, X_*) - K(X_*, X)[K(X, X) + \sigma_n^2 I]^{-1}K(X, X_*)\end{aligned} \tag{6.7}$$

In order to apply the GP framework in practice, we need to be able to train the model using a non-informative prior and a dataset with training data. By appropriate choice of the mean and covariance functions, the GP can discover the structure in the data [Do, 2007]. In the example in Fig. 6.1, we specified that the prior mean function is zero and used a kernel[5] which favors smooth functions. Namely, we used the squared exponential (SE) kernel, which is with its noise term (omitted in the example) defined as follows

$$k(x, x') = \sigma^2 \exp\left(-\frac{(x - x')^2}{l^2}\right) + \sigma_n^2 \delta_{ii'} \tag{6.8}$$

where l^2, σ^2, and σ_n^2 are the hyperparameters θ: the characteristic lenghtscale l determining

[5] A covariance function can be also called a kernel.

informally the distance in the input space related to the output values correlation[6], and the signal variance σ^2 of the output, and the variance σ_n^2 of the noise, where $\delta_{ii'} = 1$ if and only if $i = i'$ is the Kronecker's delta.

In order to make inferences about all of the hyperparameters, the probability of the data given the hyperparameters has to be computed. The distribution of the data is Gaussian and called log marginal likelihood (log evidence)

$$\log p(\mathbf{y}|X) = \underbrace{-\frac{1}{2}\mathbf{y}^T(K + \sigma_n^2 I)^{-1}\mathbf{y}}_{\text{data-fit term}} \underbrace{-\frac{1}{2}\log|K + \sigma_n^2 I|}_{\text{complexity term}} - \frac{n}{2}\log 2\pi \tag{6.9}$$

which may not be analytically tractable, but can be solved by evidence maximization

$$\hat{\theta} \in \arg\max_{\theta} \log p(\mathbf{y}|X) \tag{6.10}$$

using numerical optimization methods to find good hyperparameter settings $\hat{\theta}$ [Rasmussen, 2004]. The procedure of finding hyperparameters using evidence maximization is called training [Deisenroth, 2010, p. 15]. Since the log marginal likelihood consists of three terms: the data-fit measure, the complexity term, and the normalization constant, the evidence maximization creates an automatic trade-off, avoids overfitting and finds hyperparameters that explains the data by the simplest model. Great advantage of this training framework is that it chooses the model directly from the training data and does not require cross validation [Deisenroth, 2010, p. 15]. Other advantages can be introduced by

- exploiting different mean and covariance functions to obtain functions favoring different smoothness and length scales, introducing periodicity or non-stationary behavior [Paciorek and Schervish, 2004];
- building compositionally custom kernels via linear combination, products, integration or differentiation [Duvenaud et al., 2013];
- automatic relevance determination (ARD), which can be part of the kernel and supplements feature selection in the GP training by optimizing the characteristic length scales automatically (i.e., different weights for different input dimensions) [Deisenroth, 2010, p. 22].

One of the major drawbacks or limitations originates in the computational cost of the GP training. Training a GP model with n–sized training set using the gradient-based evidence maxi-

[6]Values further apart more than l in the input space become uncorrelated in the output space and produce less reliable regression.

mization requires $O(n^3)$ computations per gradient step (considering standard methods of inverting positive definite symmetric matrices) and memory complexity of $O(n^2)$. This computational complexity is determined by the need of kernel matrix $(K)_{n\times n}$ inversion [Rasmussen and Williams, 2006, p. 114]. This issue forms currently the main limiting factor of practical application of the standard GP. Therefore, datasets feasible on a desktop computer can include only few thousands samples [Rasmussen, 2004]. Moreover, in real applications, the training data may grow in time; thus, recomputing the whole dataset and updating the model includes the expensive matrix inversion with each new observation. It is possible to reformulate the matrix inversion as the sequential Cholesky decomposition, hence the entire matrix inversion does not have to be recalculated every time [Smith et al., 2014]. Other methods of on-line GP training may lower the need of keeping all data in memory, considering for instance a maximum number of relevant training data [Csató and Opper, 2001].

The idea of working with a reduced set of training samples leads to sparse GP approximations based on a small set of m points (where $m << n$). These m points represent a subset of original data (computational load of $O(m^3)$) or inducing points u representing a set of fictitious training data with size equal to m. Techniques using inducing points reduce the computational complexity to the computational load of $O(nm^2)$. Various sparse approximations for large problems (e.g. more than few thousand samples) for the GP regression are reviewed in [Quiñonero-Candela and Rasmussen, 2005]. Mentioned techniques include for instance a subset of data (e.g., randomly selected subset of size m), a local GP regression (i.e., training k GP models for k clusters of data and assigning test input to the closest cluster), or kernel approximations based on the inducing points u [Chalupka et al., 2013]. One of the representatives is the fully independent training conditional (FITC) approximation, which is considered as a state-of-the-art and baseline technique [Bodesheim et al., 2013]. Perhaps the most recent and promising work, that addresses scaling of the GP to large data is referred to as the distributed GP [Deisenroth and Ng, 2015]. The distributed GP are presented as conceptually straightforward, with the key idea of *recursively distributing computations to independent computational units and, subsequently, recombining them to form an overall result.* This implementation is reported to be less subject to end up in the local optima during the hyper-parameter optimization. When compared to sparse implementations, distributed GP perform the training on datasets with 10^7 data points in order of few hours.

6.2.3 Gaussian process classification

Classification is next to regression one of the important problems in machine learning. The goal is to find mapping, which is able to predict class labels for new data given labeled training data $\mathcal{D} = \{(\mathbf{x}_i, y_i)|i = 1, \ldots, n\}$ with C class labels $y_i \in \{c_1, \ldots, c_C\}$. Specifically the task is to predict the probability (i.e., probabilistic classification) that an example with input vector $\mathbf{x}$ belongs to

some discrete class. Therefore the probabilistic output lies in the interval $[0, 1]$. Further, we focus mainly on two-class classification ($C = 2$) with binary class labels $y_i \in \{-1, +1\}$.

Classification using the GP can be approached through regression. Since regression output is not restricted to the interval $[0, 1]$, there are several approaches of transforming the regression into the discrete classes. One way of relating the real-valued output of regression to discrete classes is to restrict the regression output by applying a response function. A common choice is the logistic function $\lambda(z) = (1 + \exp -z)^{-1}$ (i.e., sigmoid function[7]) [Rasmussen and Williams, 2006, p. 4]. Placing sigmoid over the prior function $f(\mathbf{x})$ results in class probability $\pi(\mathbf{x} = p(y = +1|\mathbf{x}) = \sigma(f(\mathbf{x})))$. This is illustrated in Fig. 6.2 by a one dimensional example. Once the GP prior is transformed using the sigmoid, there is no straightforward analytical solution for the posterior, since it involves an integral that is the product of a Gaussian and sigmoids.

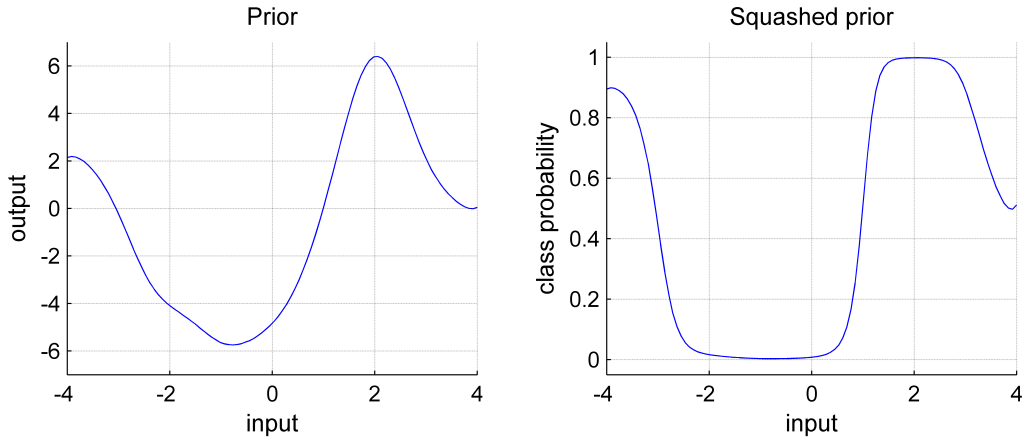

Figure 6.2: Gaussian Process classification - prior and squashed prior. Left plot shows one function randomly sampled from the GP prior. Right plot shows this function squashed using the logistic function (sigmoid). Class probability is reduced to interval $[0, 1]$.

Thus, the analytic approximations of integrals, which approximate the non-Gaussian joint posterior with a Gaussian one, have to be used. Such approximation methods include for instance Laplace approximation (LA), expectation propagation (EP), or computationally demanding Markov chain Monte Carlo sampling (MCMC) [Rasmussen and Williams, 2006, p. 40]. The Laplace's method uses a second order Taylor approximation of the log posterior to find the approximation [Williams and Barber, 1998]. The EP approximation is an extension of the density filtering incorporating iterative refinement of the approximations [Minka, 2001]. In principle, MCMC sampling becomes exact in the limit of long runs; therefore, it can be used as the standard by which the comparison of the LA and EP approximation methods are evaluated [Kuss

[7]A sigmoid function is a bounded function, which has positive derivative at each point; it is shaped like a letter S.

and Rasmussen, 2005]. Despite it is strongly suggested to choose EP for approximate inference in binary models [Kuss and Rasmussen, 2005], it is also mentioned, that evaluation of the approximations as functions of hyperparameters cannot be generally done. Therefore, an empirical assessment for a specific dataset reveals practical usefulness of the approach.

6.2.4 Anomaly detection with Gaussian processes

Gaussian processes have been proved beneficial in many applications in diverse fields, where the underlying process of interest is intended to be recovered from data. Anomaly detection[8] is not an exception and it has been reported, that the GP outperform the state-of-the-art methods even in this field. Some of the areas, where the GP are successfully applied in anomaly, novelty, or outlier detection, cover visual object recognition [Kemmler et al., 2013, Ramirez-Padron et al., 2013, Bodesheim et al., 2013], multi-class image recognition [Gao and Li, 2011], attribute prediction, defect localization and background subtraction in video sequences [Kemmler et al., 2013], removal of drift and bias errors in faulty sensor data [Reece et al., 2009], detecting suspicious behavior of maritime vessels [Smith et al., 2014, Kowalska and Peel, 2012] or fault detection in process simulations [Serradilla et al., 2011].

In our analysis of suitability of the GP for anomaly detection within a Kalman filter framework, we focus on several different and we propose following approaches or options for further investigation.

6.2.4.1 GP anomaly detection approaches

There are several ways of dealing with the problem of detecting anomalous measurements using the GP. The first two choices are naturally represented by the GP regression and classification (see Section 6.2.2 and 6.2.3). It has been pointed out, that the GP regression (see Section 6.2.2) can be used to model the normal behavior and derive the anomalies as samples outside this normal area [Kemmler et al., 2013,Bodesheim et al., 2013]. This one-class setup is particularly important, when it is either very expensive or impossible to collect sufficient set of training samples [Ramirez-Padron et al., 2013]. Generally, there is rarely enough data available to create a representation, that covers the whole feature space and provides perfect generalization. There is, however, a significant drawback in the one-class setup—automatic training of hyperparameters fails in the one-class setting due to the absence of a trade-off between data fit and model complexity [Kemmler et al., 2013]. Naturally, when dealing with a rich and well-balanced dataset, regression or binary

[8] Anomaly, novelty, abnormality, outlier—all these terms are being used in different works to describe patterns within the data, which do not conform to expected behavior; the system is not aware of them during some process; this terminology may be in contradiction with our definition of the anomalous data in Chapter 5; however, we preserve our initial terms, where anomalous measurement is an observation, that causes significant deterioration of the state estimates.

classification is well conditioned by the data and the single class can be represented by either normal data or data labeled as anomalous. In our specific dataset, we have about 2% of anomalous data (see Chapter 5), therefore another aspect to examine are be the ratios of the two classes.

6.2.4.2 Proposed approaches

We propose following approaches for further investigation in the anomaly detection problem: GP classification (GPC), normal class GP regression and anomalous class GP regression. Selected approaches are be examined for their performance with unbalanced and balanced datasets, and compared with the sparse implementations, which are be necessary for speeding up the training process.

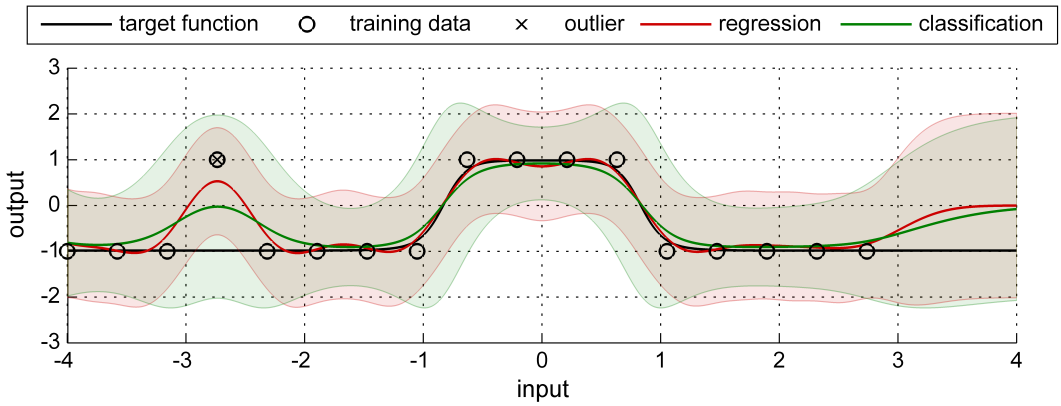

Figure 6.3: Binary classification illustration example - GP regression (red) and GP classification (green); both trained with SE kernel and zero mean function. Training data representing classes $\{-1, +1\}$ are sampled according to the target function (black) and marked as circles. One demonstrative outlier is marked by a cross. Test data are evenly spaced in the input interval $[-4, 4]$, transparent areas corresponds to the uncertainty interval.

The difference between binary classification done by GP regression (GPR) and GP classification is illustrated in Fig. 6.3. This explanatory 1D example shows, how these two approaches behave on a binary classification problem and how the predictive output mean and covariance evolves with respect to the training data. Both GPR and GPC handle this simple problem well when trained with the automatic optimization of the SE kernel hyperparameters. The example also shows, how both approaches behave, when receiving artificially introduced outlier and missing samples in the training data. In this specific example, a single wrong training sample (denoted as outlier) is handled better by the GPC (i.e., predictive mean is shaped such as it prefers the right class and the variance is predicted accordingly). In the case of missing data, the predictive output is, as expected, attracted to the zero mean prior in both approaches.

Explanatory instance of the one-class regression method is shown in Fig. 6.4. We illustrate both options - GPR trained with normal data and with anomalous data. Both approaches exploit natural behavior of the GP—predictive output is returning to the initial belief (GP prior) where the input space was not covered during the training. The GP prior is set, such as it represents the opposite of the single class used for training. As discussed, main drawback of this approach is the impossibility of automatic tuning of the hyperparameters. Solvers maximizing the evidence (6.10) run toward extreme cases (i.e., completely flat or extremely responsive predictive mean and unrealistic predictive variance) or crash due to numerical instabilities [Kemmler et al., 2013]. Tunning of hyperparameters can be done in a grid-search cross-validation manner, when the other class is used to specify the performance under a particular hyperparameter setup. Other remedies for treating tuning of parameters may embrace centering the data for usage of the zero mean function, standardize the input data (very common pre-processing in machine learning) and class targets, and setting the lenghtscales and variance to one, respectively [Murray, 2008].

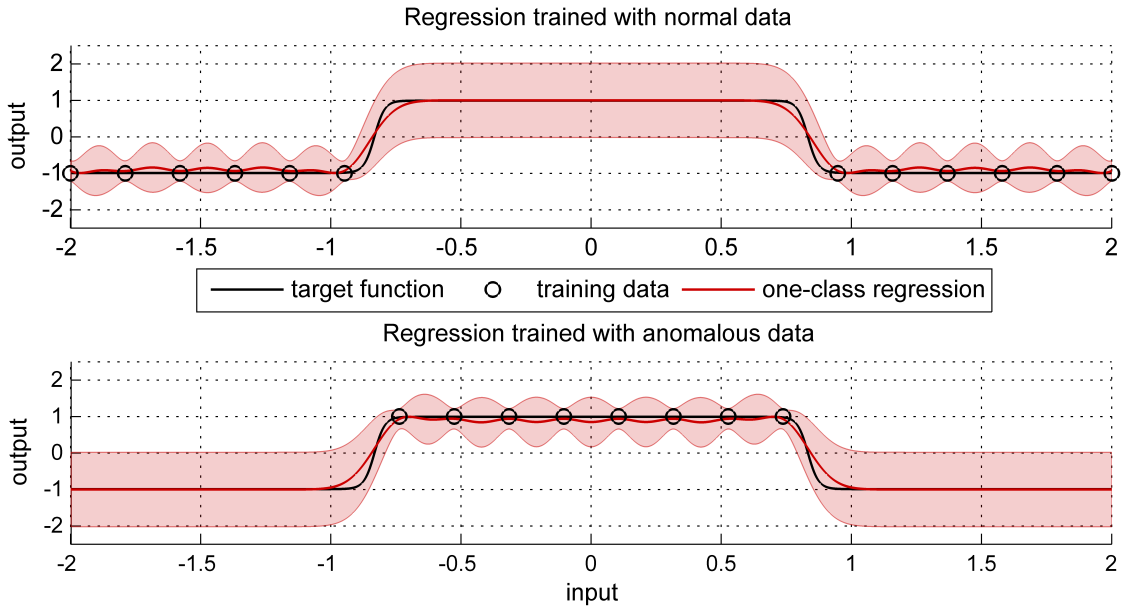

Figure 6.4: Binary classification illustration example - one-class GP regression (red); trained with SE kernel ($l = \log 0.1, \sigma^2 = \log 1$). Training data representing classes $\{-1, +1\}$ are sampled according to the target function (black) and marked as circles. Top figure represents regression trained only on normal data (class $\{-1\}$), bottom figure represents regression trained only on anomalous data (class $\{+1\}$). Test data are evenly spaced in the input interval $[-2, 2]$, transparent areas corresponds to the uncertainty interval.

Naturally, selection of the mean and covariance functions determines the behavior of the predicted output. As shown in previous two examples, different mean functions, even the constant ones, can have large impact on the results and in some cases they can involve great portion of

our belief in the underlying process. We don't expect the mean function anomalies to have linear or polynomial character, therefore we remain with the constant and zero mean functions.

There is a large number of options for the covariance functions, even custom ones. Most common choice among multiple applications is the squared exponential (SE) kernel [Kemmler et al., 2013, Smith et al., 2014, Kowalska and Peel, 2012, Zimmermann et al., 2015], which is often supported by the automatic relevance determination (ARD). Prediction generated using the SE kernel resemble natural and smooth functions. Another popular kernels are the rational quadratic covariance function, which can be seen as an infinite sum of SE kernels with different lengthscales, and the Matérn covariance [Smith et al., 2014], which depends only on the Euclidean distances of the points and its parameter directly controls the level of smoothness [Rasmussen, 2006]. We demonstrate the differences between the SE, Matérn $1/2, 3/2$ and $5/2$, and RQ kernels in Fig. 6.5 (all kernels are implemented with the ARD, even though it does not have any effect in this specific 1D example). It is clearly evident, that for this demonstrative example, the RQ is close to the SE (for a specific settings of the RQ hyperparameters, the RQ is identical to the SE kernel). Downside of the smooth covariance functions (e.g., SE and RQ) is that they struggle with discontinuities in the data. If there is such discontinuity, smooth kernels end up with lengthscales optimized according to the smallest wiggle in the function, resulting in a wrong prediction in the smooth regions (lengthscales are short and posterior mean forced to be close to prior as a consequence) [Duvenaud, 2014].

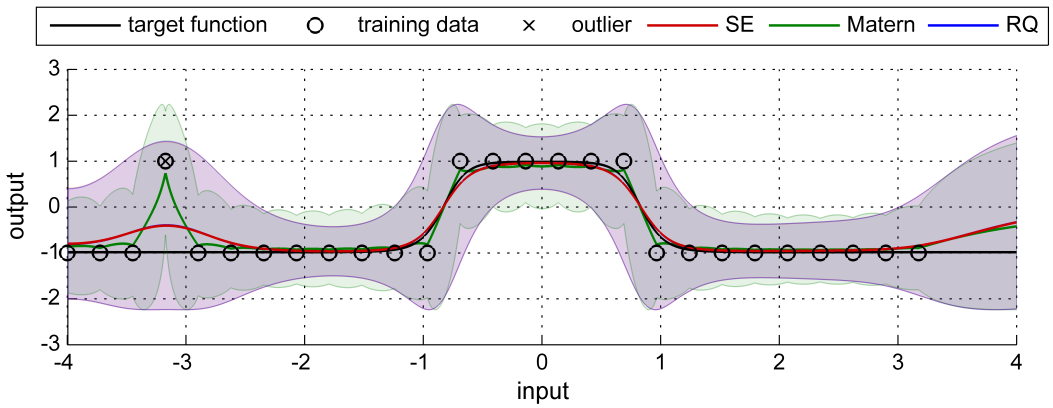

Figure 6.5: Binary classification example - GP classification with different covariance functions. Training data representing classes $\{-1, +1\}$ are sampled according to the target function (black) and marked as circles. GP classification trained with following kernels: squared exponential (red), Matérn $1/2$ (green), rational quadratic (blue). One demonstrative outlier is marked by a cross. Test data are evenly spaced in the input interval $[-4, 4]$, transparent areas corresponds to the uncertainty interval.

6.3 Experimental evaluation

In this section, the goal is to present both elaborate assessment of the anomaly detection performance as well as comments and discussion of the analyzes. This experimental evaluation consists of following parts, which are all related to the dataset and experiment evaluation done in Chapter 5 and we consider following results as a continuation of this work. First, in Section 6.3.1 we test the proposed approaches from Section 6.2.4 with our anomaly detection dataset presented in Chapter 6. Second, in Section 6.3.2 we provide the comparison of the selected GP approaches with respect to the methods analyzed in Chapter 6—covariance (COV), chi-squared (CHI), Gaussian Mixture Models (GMM), one-class Support Vector Machines (SVM), and logistic regression (REG). Third, in Section 6.3.3 we apply the selected GP approach in the four test case environments (indoor, outdoor urban, outdoor disaster site, outdoor forest). Finally, we provide discussion, comments and thoughts on using GP in the anomaly detection framework for multi-modal data fusion. Since the evaluation include thorough search for the best GP model for our anomaly detection problem, all the performance tables are placed in Appendix A.

6.3.1 Classification performance

In this section, we evaluate the classification performance of the proposed GP approaches in the view of several different aspects: selection of the regression or the classification, sparse implementation, selection of the mean and the covariance functions, and utilization of the uncertainty predictions. For each aspect, we describe in what parameter space we searched for the best performing settings, and provide detailed results by the means of tables describing classification performance in the three exteroceptive modalities: visual (attitude) and laser (attitude and position). For completeness we restate that our publicly available classification dataset used for training and testing of the different classifiers has approximately 1.7 km of indoor data with precise and accurate ground truth and anomalous measurements labeled as anomalous, others as normal. It contains 20 runs, which represent standard conditions, and 25 runs with failures of the exteroceptive modalities. We evaluate the approaches in a different dataset from several challenging environments—test cases—having approximately 1.2 km of traveled distance. More details about the dataset can be found in Chapter 6. In our training scenario, the data were shuffled randomly and normalized such as each feature has zero mean and variance equal to one. If not specified otherwise, all evaluation contains following metrics: True Positives (TP), False Positives (FP), True Negatives (TN), False Negatives (FN), Precision (PR), G-mean value (G), and Area under the Receiver Operation Characteristic (ROC) Curve (AUC). All the metrics are evaluated on the testing portion if the dataset at the best classifier threshold according to the best G-mean value obtained from the training metrics.

6.3.1.1 Classification and regression

In the comparison of GP classification and one-class regression, the gradient-based hyper-parameter optimization did find a solution for both of the methods. However, the resulting one-class regression models predicted only one class, no matter what inputs were given. This is a behavior described in the literature [Kemmler et al., 2013]. We never encountered such problems with the GP classification and the hyper-parameter optimization always converged in our training dataset. Therefore, we decided to pursue solely the GP classification trained with both anomalous and normal data in our further analyses.

Sparse implementation We carried out a thorough evaluation of two sparse implementations: approximation subsampled data (i.e., randomly or evenly selected subset of data) and FITC approximation. All experiments were evaluated using the GP classification with zero mean function and squared exponential kernel with automatic relevance determination.

1. **Subsampled data**
   Selecting a subset of the data is one of the most straightforward approaches to sparse implementations of GP. We study the influence of different features and the number of negative samples. Number of positive samples is fixed and equals to the maximum number of positives available in our dataset. Therefore, we compare performance on following settings:

   - features (norm of innovations or innovation vector)
   - balanced data (the number of positive samples equals to the number of negative samples)
   - $10, 30, 50, 100$ % of all negative samples

   We list the results in Tab. A.1, A.2, A.3 in Appendix A.

2. **FITC approximation**
   The FITC approximation by inducing points is considered as the baseline technique for sparse implementations. In this evaluation, we study following aspects:

   - features (norm of innovations or innovation vector)
   - number of inducing points
   - range of inducing points (fixed or defined by minimum and maximum of the features)
   - distribution of inducing points (equispaced or random)

   We list the results in Tab. A.4, A.5, A.6 in Appendix A.

Mean and covariance functions In the search of the best mean and covariance functions, we follow the common practice and set the mean function to be constant and zero. The search therefore includes only different covariance functions:

- features (norm of innovations or innovation vector)
- covariance functions (SE, RQ, Matern $1/2, 3/2, 5/2$)

We list the results in Tab. A.7, A.8, A.9 in Appendix A.

Uncertainty of predictions As the GP directly outputs the uncertainty in terms of predicted variance, we study the possibility of discarding the predictions, which have high uncertainty. One scenario was tested for different level of the uncertainty discarded:

- uncertain predictions with uncertainty greater than certain threshold are classified as negatives

We list the results in Tab. A.10, A.11, A.12 in Appendix A.

6.3.1.2 Summary and discussion of the classification performance

Purpose of this section is to provide summary of the classification performance and discuss the results. We assess the individual evaluations in the order they have been presented.

Sparse implementation - subsampled data Main reason to use sparse implementations is to relief the computational burden, because especially data with more than few thousands samples are difficult handle in reasonable time. In our case, the processing time for subsampled data spanned from few second in the case of the balanced data, to several hours for the whole dataset (i.e., 6 hours for VO attitude). It is evident, that it is beneficial to use more negatives, presumably balanced data don't provide enough samples to cover the whole feature space. There is no direct indication, that using all the samples increase the performance the most (best results were obtained for 50% of negatives, see Tab. A.1 model #8 for VO attitude, Tab. A.2 model #7 for laser attitude, and Tab. A.3 model #8 for laser position). Results with balanced sets have more FP, but the difference is not that crucial, especially if the application aims for really fast training phase.

Sparse implementations - FITC approximation As expected, the FITC implementation performed in some parameter settings slightly better than the straightforward subsampling. The other parameters searched, such as number of inducing points, their range and distribution did not have very large effect on the final results. The best results for all modalities were reached

by the model #19 with equidistant distribution of 5 inducing points over the whole feature space (see Tab. A.4, A.5, and A.6). It is worth mentioning, that there are models, where the training failed and predicted only one class. Therefore, we claim it is necessary to search the parameter space for the best settings for a proper deployment.

Mean and covariance functions Once more, the necessity of searching best parameters repeats and the choice of covariance function shows to be very application specific. For instance, in the case of the VO attitude, the SE and RQ kernels outperform the Matern kernels (see Tab. A.7 models #2 and #4), however, in the case of laser modalities, Matern kernel performs similarly (see Tab. A.8 and A.9). Thereby we confirm, that the smooth SE kernel with automatic relevance determination, a frequent choice among researchers, is also very viable option for our anomaly detection problem.

Uncertainty of predictions Estimating uncertainty of predictions is often discussed in works implementing the GP [Santamaria-Navarro et al., 2015] [Kemmler et al., 2013]. Having the prediction as well as its uncertainty definitely helps to distinguish the cases, which are too far from the training set and have large uncertainty. Our proposed treatment of uncertain predictions (i.e., assigning too uncertain predictions to the negative class) decreased the number of FP, therefore increased precision. The threshold was always set to reflect the percentage of the maximum estimated uncertainty. Most noticeable results were observed in the case of VO modality, where the precision increased by about 70% for a threshold limit of 80% that retained 96% of valid data (model #3 in Tab. A.10). However, it was also shown, that the threshold limit is very application specific and has to be tuned to match the performance expectations. For instance setting the threshold as low as 50% can retain 91% of the VO modality data, but only 12% of the laser data, which is in no case desirable outcome. Unfortunately, when we tried to discard uncertain predictions in the best models, we obtained so far (model #19 in Tab. A.4, Tab. A.5, Tab. A.6), it resulted in lowering the precision and G-mean values, since the threshold discarded too many TP in the process. In the end, using uncertainty as other information source can be beneficial in some cases; however, we have shown, that it does not provide any other performance improvements in an already fine-tuned model.

6.3.2 Comparison with other anomaly detection methods

In this section, we compare the results achieved with the selected GP model (model #19 in Tab. A.4, Tab. A.5, Tab. A.6), with the classification methods studied in Chapter 5. We introduce another metric, the receiver operating characteristic (ROC) curve, which is a popular metric for comparing binary classifiers. Furthermore, we update the classification performance table from

Chapter 5, to show how the GP models stand against other methods in more detailed look.

The ROC curve shows the ability of the classifier to rank the positive instances relative to the negative instances [Fawcett, 2004]. By plotting the true positive rate (TPR) with respect to the false positive rate (FPR), the ROC depict in 2D the trade-off between hit rates and false alarm rates of classifiers. The whole ROC curve is generated by adjusting the threshold of score or probability output of the classifier. In the end, such curve evaluated on training set may help to set the score threshold to achieve desired values of TPR and FPR. Common way to reduce this 2D metric into a scalar is to compute the area under the curve (AUC). The AUC values are always between 0 and 1, where the perfect value is the area of 1.0 and random guessing is represented by the area of 0.5.

In Fig. 6.6, we present the ROC curves and AUC values for all the evaluated classifiers from Chapter 5 and Chapter 6. This evaluation follows the result from Chapter 5 with few updates. First, we add also the GP model in the comparison. According to the evaluation done in Section 6.3.1.2, the GP model scores the expected performance of the highest performing classifiers represented by the GMM and the one-class SVM. Second, we revisit the performance evaluation of the logistic regression classifier (REG), which was not performing satisfactory in the results presented earlier. It was shown, that the regression learns the underlying process of anomaly detection very well. However, the probability threshold must be further tuned on the training set to obtain the best performance. Such assessment was actually brought by the introduction of ROC and AUC metrics, which provide another point of view at the classification performance.

One of the advantages of the ROC curve is that it is unaffected by the classes skew in assessing the performance of unbalanced datasets [Jeni et al., 2013]. The problem is however apparent for highly unbalanced data, where the differences between curves tend to be very small, but still present. Therefore, it is necessary to measure the performance also by other means, since the ROC may mask poor performance. It can be clearly seen from Fig. 6.6 that the AUC values of all machine learning classifiers perform substantially better that the standard covariance and chi-squared gating tests. At the same time, the effect of highly unbalanced data can be observed in the small differences between highest scoring classifiers (GMM, SVM, REG, GP).

To complete the performance results, in Table 6.1 we update the classification performance table from Chapter 6. Since TPR, TNR and G-mean values may mask poor discrimination of the TP and FP, we also add the precision metrics $\mathrm{PR} = \mathrm{TP}/(\mathrm{TP} + \mathrm{FP})$ to correctly address and discriminate such problem. As in the evaluation using the ROC, we update the confusion matrix metrics in the case of logistic regression (REG) in Table 6.1. It turns out, that by finding the right score threshold in the training set, even logistic regression is capable of achieving performance of the highest scoring classifiers. However, in all evaluated modalities, precision of the REG is lower than for instance that of the GMM or GP. In the end, by looking at the two scalar metrics,

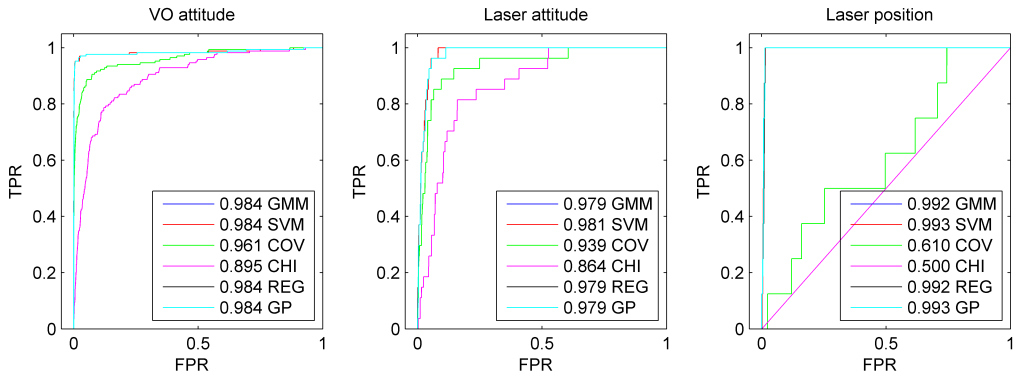

Figure 6.6: Receiver Operating Curve and Area Under the Curve as obtained on the testing examples. True Positive Rate (TPR) versus False Positive Rate (FPR) for all three modalities. AUC values are on the left of classifier abbreviations.

G-mean and precision, we can identify the most successful classifiers for each modality—GP for the VO attitude and laser attitude modalities, and GMM for the laser position modality.

6.3.3 Localization performance

In this section, we follow the application of detection and rejection of anomalous exteroceptive data in field test cases presented in Chapter 5. We verify the anomaly detection with the GP classifier, to show, how the classification results reflect in real improvement in the field testing. Therefore, the evaluation consist of the results without any anomaly detection, results with the best anomaly detection (AD) from Chapter 5 (GMM classifier), and results with the anomaly detection done by the GP model. Namely, we show the results for four test cases: indoor experiment with the Vicon ground truth, outdoor urban experiment with the Leica ground truth, disaster training site, and outdoor forest experiment with the GPS ground truth. All these test cases are introduced and described in Chapter 5 in detail.

1. Indoor Vicon room

As can be seen from Fig. 6.7 GMM and GP anomaly detections yield identical results. This is an expected result. In this particular indoor experiment, the robot encountered few major outliers in the VO attitude, and GMM and GP handled those without any issues.

2. Leica experiments

Results from the four different experiments carried out in an urban environment are depicted in Fig. 6.8. Especially the bottom part of the graph shows, that the GP performed very similarly as the GMM and even better in the first, second, and fourth experiment when comparing the average position error, and better in the final position error (i.e., the last sample in the average position error) in all experiments.

Table 6.1: Classification results as obtained on the testing examples

	Features	TP	TN	FP	FN	TPR	TNR	PR	G
VO attitude									
COV	$\Delta\mathbf{x}_{\mathtt{att}}$, $\sigma_{\mathtt{att}}$	157	6773	971	12	0.93	0.87	0.139	0.901
CHI	$\mathbf{v}_{\mathtt{att}}, \mathbf{S}_{\mathtt{att}}$	130	6867	877	39	0.77	0.89	0.129	0.826
GMM	$\|\mathbf{v}_{\mathtt{att}}\|$	160	7702	42	9	0.95	0.99	0.792	0.970
SVM	$\mathbf{v}_{\mathtt{att}}$	160	7701	43	9	0.95	0.99	0.788	0.970
REG	$\|\mathbf{v}_{\mathtt{att}}\|$	161	7690	54	8	0.95	0.99	0.748	0.973
GP	$\|\mathbf{v}_{\mathtt{att}}\|$	160	7705	39	9	0.95	0.99	**0.804**	**0.971**
Laser attitude									
COV	$\Delta\mathbf{x}_{\mathtt{att}}$, $\sigma_{\mathtt{att}}$	26	1170	407	1	0.96	0.74	0.060	0.845
CHI	$\mathbf{v}_{\mathtt{att}}, \mathbf{S}_{\mathtt{att}}$	14	1429	148	13	0.52	0.91	0.086	0.685
GMM	$\|\mathbf{v}_{\mathtt{att}}\|$	25	1499	78	2	0.93	0.95	0.242	0.938
SVM	$\mathbf{v}_{\mathtt{att}}$	26	1473	104	1	0.96	0.93	0.200	0.948
REG	$\|\mathbf{v}_{\mathtt{att}}\|$	26	1471	106	1	0.96	0.93	0.197	0.948
GP	$\|\mathbf{v}_{\mathtt{att}}\|$	25	1505	72	2	0.93	0.95	**0.257**	**0.940**
Laser position									
COV	$\Delta\mathbf{x}_{\mathtt{pos}}$, $\sigma_{\mathtt{pos}}$	0	1596	0	8	0.00	1.00	NaN	0.000
CHI	$\mathbf{v}_{\mathtt{pos}}$, $\mathbf{S}_{\mathtt{pos}}$	0	1596	0	8	0.00	1.00	NaN	0.000
GMM	$\|\mathbf{v}_{\mathtt{pos}}\|$	8	1574	22	0	1.00	0.99	**0.266**	**0.993**
SVM	$\mathbf{v}_{\mathtt{pos}}$	8	1488	108	0	1.00	0.93	0.068	0.965
REG	$\|\mathbf{v}_{\mathtt{pos}}\|$	8	1518	78	0	1.00	0.95	0.093	0.975
GP	$\|\mathbf{v}_{\mathtt{att}}\|$	8	1507	89	0	1.00	0.94	0.082	0.972

Note: $\Delta\mathbf{x}$ is the EKF error state vector, $\mathbf{v}$ is the EKF residual vector, $\mathbf{S}$ is the EKF residual covariance matrix, σ is the EKF standard deviation; $\|\mathbf{x}\|$ denotes Euclidean norm of a vector $\mathbf{x}$; subscripts `att` and `pos` denote attitude and position elements in a vector or matrix.
TP – True Positives; TN – True Negatives; FP – False Positives; FN – False Negatives; TPR – True Positive Rate; TNR – True Negative Rate; PR – Precision;G – G-mean value.
COV – Covariance analysis; CHI – Chi-squared test; GMM – Gaussian Mixture Models with 2 mixtures; SVM – One-class SVM with RBF kernel; REG – Logistic Regression; GP – Gaussian Process.

3. Prato disaster site

It is evident from Fig. 6.9, that even in this case, the GP performed competitively with the GMM in the environment of a site simulating disaster conditions. The robot always started and ended at the same position in these experiments. A closer look at the second experiment (Fig. 6.9 - right) reveals, that the GP trajectory finishes closer to the starting point. However, due to the complex environment of the training disaster site, the ground truth is not available in these experiments.

4. Outdoor forest

The results from the final experiment, the most challenging one situated in a forest, are depicted in Fig. 6.10. We can observe, that the GP anomaly detection (green dashed line) performed better than GMM (blue line) during the first half of the trajectory—the GP trajectory is much closer to the GPS trajectory (black dash-dot line). For the majority of the experiment, the GP approach ensures that the trajectory is within the bounds of the ground truth from the GPS (gray area). The second half of the experiment, especially the final position is however worse than in the case

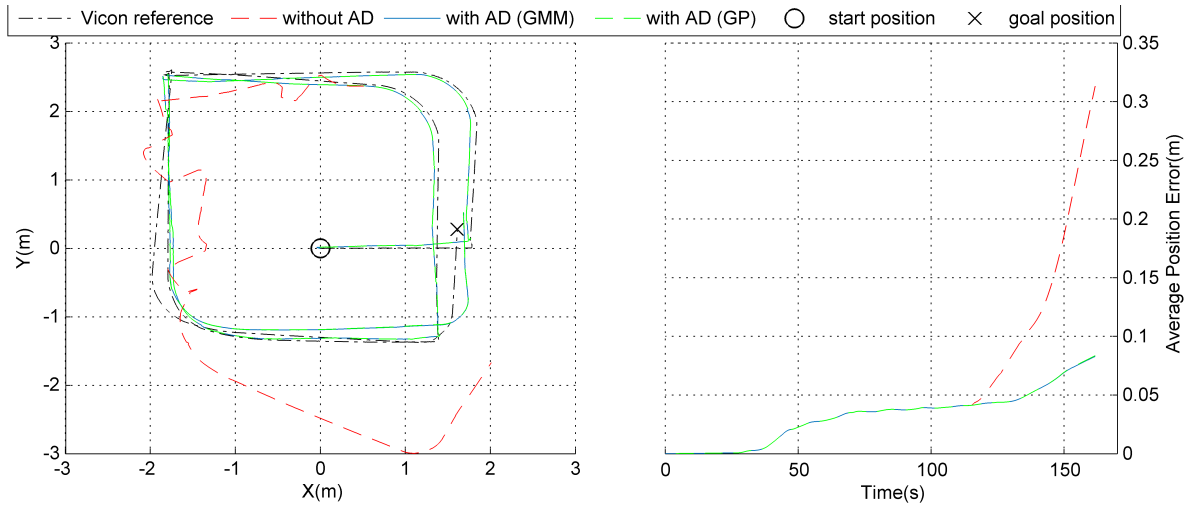

Figure 6.7: Indoor experiment - position estimates with GP and GMM anomaly detection. (Left) Shows 2D projection of the trajectories. (Right) Average position error expressed as function of time. VO attitude failure between 115 s and 140 s. Black dash-dot - Vicon reference; red dashed - data fusion without AD; blue - data fusion with AD (GMM); green - data fusion with AD (GP).

of the GMM approach. This is caused by the nature of the dead-reckoning—even a slightest error in the yaw angle propagates through the rest of the experiment and significantly influences especially the estimated position. We can see, that this yaw error most probably occurred during multiple turns shown in the lower left part of the upper plot of the figure. From there, the position estimate of the GP approach develops almost identically as for the GMM approach. The overall average position error is in total smaller for the GP approach, which is another confirmation of the results presented in the classification performance section.

6.3.4 Discussion of the GP performance

The findings and results presented in Chapter 6 are an extension of the anomaly detection framework introduced in Chapter 5. In this discussion, we emphasize findings regarding GP classification and general findings regarding the anomaly detection framework. First of all, we verified that the popular choice of squared exponential kernel with the zero mean function is a legitimate choice even for the anomaly detection problem. In the search for the best possible parameters, we often reached similar results when discovering the parameter space. However, in the end we claim, that it is worth to optimize the parameters and method settings by a grid search or by optimizing methods searching for the best performance of the classifier. We have also revisited the results from Chapter 5 and introduced the ROC and AUC metrics. Both helped us to tune

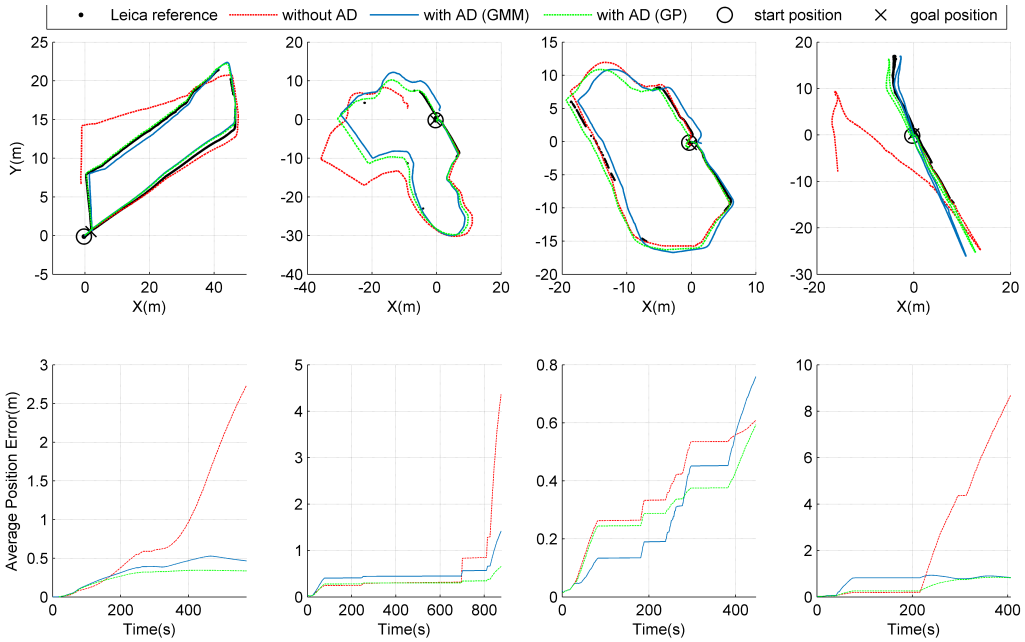

Figure 6.8: Urban environment - position estimates with GP and GMM anomaly detection. (Top) 2D projections of the trajectories. (Bottom) Average position errors expressed as function of time. Black - Leica reference; red - data fusion with AD (GMM); green - data fusion with AD (GP).

and evaluate the performances of the classifiers. The threshold in the ROC, represented by score (i.e., classification probability) generated by the classifier, may be tuned by the needs of the final application—the performance is always a compromise between the true and false classifier detections. Despite there were only minor improvements in the GP classification performance, with respect to results presented in Chapter 5, we registered an improvement in the localization performance (see urban and forest experiments in Fig. 6.8 and 6.10). These results directly imply that improving classification performance on the training dataset also improves the localization performance on unseen experiments.

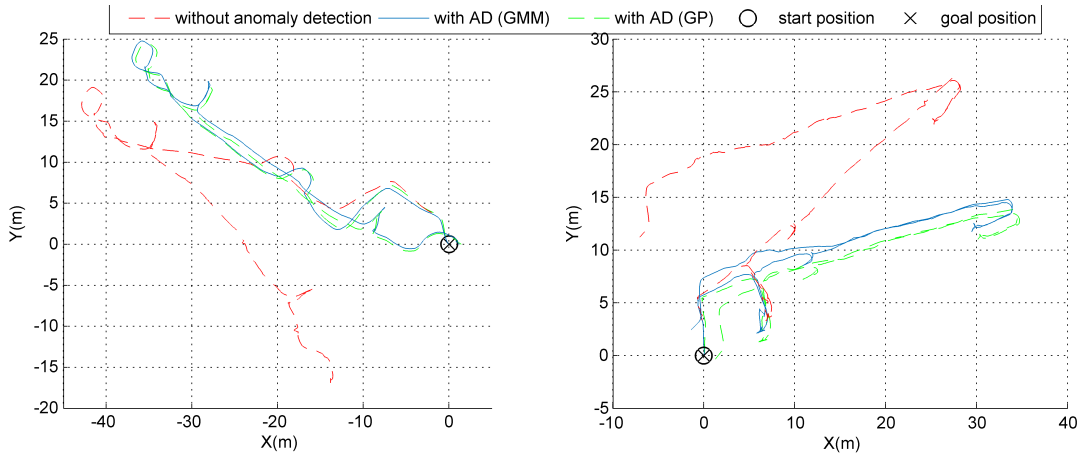

Figure 6.9: Prato training disaster site - position estimates with GP and GMM anomaly detection. (Left) Disaster site 1. (Right) Disaster site 2. Black - Leica reference; red - data fusion with AD (GMM); green - data fusion with AD (GP) trajectories.

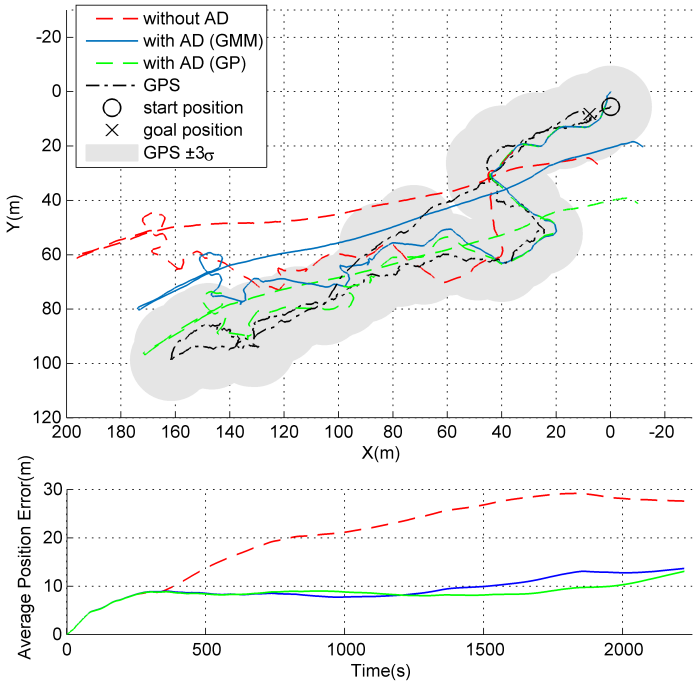

Figure 6.10: Forest experiment - position estimates and error plots with GP and GMM anomaly detection. (Top) 2D projection of the trajectories. (Bottom) Average position error expressed as function of time. Black dash-dot - GPS; red dashed - data fusion without AD; blue - data fusion with AD (GMM); green - data fusion with AD (GP); gray area - experimentally determined GPS uncertainty.

Chapter 7

Conclusions

7.1 Summary of the thesis and contributions

In this doctoral thesis we examined localization of a vehicle represented by a skid-steer mobile robot. We approached the localization by the methods of state estimation supported by machine learning.

In the first part of the thesis, we showed in Chapter 4, that a ground vehicle is capable of having computationally not demanding and reliable dead-reckoning system based solely on inertial measurement unit and velocity sensor placed on wheels or tracks. We did go beyond standard experimental evaluation and utilized the core IMU and odometry dead-reckoning in real search-and-rescue missions. The skid-steer robot was deployed in the danger zone to assess earthquake damage to historical buildings and cultural artifacts located in Mirandola, Italy [Kruijff et al., 2012]. The teleoperated robot helped to perform such tasks as exploration of a danger-zone disaster site. Our fusion of the IMU and odometry provided low-level localization system necessary for high-quality 3D mapping of the environment, and helped to enhance the situational awareness and lower the cognitive load by providing more concise knowledge about robot pose to the operator. According to our knowledge, we presented above-standard, comprehensive evaluation of the IMU and odometry dead-reckoning in various environments and specified the average position error related to the distance driven. Our major achievements in the first part of the thesis could be summarized as follows:

- Development and implementation of four estimation architectures consisting of the combination of the extended Kalman filter in both direct and indirect form and a complementary filter for attitude estimation.
- Thorough comparison of localization performance and computational demands of the four different EKF estimation architectures fusing IMU and odometry.

- Comprehensive verification and testing of the localization system in indoor and outdoor experiments with more than 4 km of distance traveled.
- Reaching the average position error of 4 % of the distance traveled (results of the best candidate).
- Identification of the best approach for the deployment in a real search and rescue mission—the EKF in the direct form combined with the complementary filter.
- Usage of the best estimation architecture on-board of a search and rescue robot, which was deployed in a real mission to assess the damage in earthquake-affected areas (see Chapter 4).

In the second part of the thesis, we proposed usage of supervised machine learning methods for monitoring and inspection of the Kalman filter innovations in Chapter 5 and Chapter 6. We implemented the anomaly detection extension of a multi-modal data fusion framework. We also proved, that using a trained classifier for anomaly detection provides better anomaly detection performance than the standard methods. We supported both of our main tasks by thorough experimental evaluation including testing the systems, IMU and odometry dead-reckoning, and the anomaly detection extension of a multi-modal data fusion, on the data gathered using a real platform under real and challenging conditions. We presented following achievements in the second part of the thesis:

- Extension of a multi-modal data fusion for localization of a mobile robot by anomaly detection in the exteroceptive measurements
- Methodology for preparing an anomaly detection training dataset with the use of accurate and precise ground truth information about the robot pose
- Implementation and training four different machine learning approaches for anomaly detection in a binary classification manner
- Comparison of classification performance of the standard tests, used in the Kalman filter monitoring, with the proposed machine learning approaches
- Thorough verification of the best anomaly detection candidates in localization in challenging indoor and outdoor environments
- Localization improvement in the final position error of distance traveled from (7.9 ± 6.1) % without anomaly detection to (1.5 ± 1.1) % with anomaly detection (mean $\pm$ standard deviation)

Our findings present an incremental step towards the autonomous navigation, exploration, or unaided assist to search-and-rescue teams. Mobile robotics advances rapidly. however, deployment in real and challenging conditions is still matter of an extensive research.

7.2 Future work

In this section, we propose several suggestions for future research extending the work presented in this doctoral thesis.

- Since the work implements the dead-reckoning, a promising chapter for future research would be to merge the relative and absolute positioning systems, and allow the robot to map its environment and update its belief of its pose. This is also being extensively researched and known as the Simultaneous Localization and Mapping problem (SLAM). There is definitely a lot of work done in the past decade in the field of robotics. However, the perfect data-fusion enabling flawless localization under all possible conditions has not been reached yet.

- The platform of interest in this thesis is represented by a ground vehicle. In this platform, the algorithms are dependent only on one particular modality specific for a ground vehicle—the odometry—and could be applied on any vehicle, wheeled, tracked, legged or aerial. The wheeled or tracked odometry can be replaced by another sensor measuring velocity in the frame of the vehicle (e.g., legged odometry or versatile visual odometry). Thus, all the algorithms are definitely transferable to any other platform equipped by a similar sensor set. Therefore, it would be interesting to see how the trained anomaly detectors perform when used on another platform. Other option for a different platform would be to prepare similar training datasets, including the fail cases for all the modalities and supported by a precise and accurate ground truth. We understand, that this can be time-consuming and tedious, however, majority of the research labs dealing with navigation already possess the needed equipment—platforms and ground truth systems. Other option would be to use the one-class classifiers trained on data without any anomalies. This has to be ensured to enclose the normal feature space correctly and to maximize the probability of finding anomalous samples. The possible extensions to other robotic platforms include:

  - aerial vehicles, which are currently based mainly on IMU and GPS fusion for outdoor applications, and IMU and camera fusion is being extensively studied by numerous research labs for indoor and local outdoor missions;
  - legged robots, which can carry almost the same sensors and the connection between a legged odometer, exteroceptive pose estimates and anomaly detection is at least as crucial as in the case of wheeled vehicles; legged robots have usually much faster dynamics than a wheeled vehicle and are considered as a 2.5D platforms (i.e., they are bounded to the ground, but they are exposed to aerial phases due to the nature of their movement patterns).

- One of the areas, that has not been researched in the anomaly detection part of the thesis is the context surrounding the decisions. Implementation of our methodology could be further improved by following:
  - introducing the time dependency in anomaly detection; every anomaly is surrounded by specific causes and consequences, which do not happen in a single time instance, as we suppose in our implementation;
  - allowing for interesting features also in the core exteroceptive algorithms—visual odometry and iterative closest point matching; these should provide another view and help the classifiers to generalize by adding more information; in our implementation the main limiter was the computational cost of adding such features.

  So far, the main reasons preventing us to pursue these improvements are the different rates at which different modalities operate. Once the sensors will allow for higher sampling frequencies, computational power will be sufficient to process the sensor outputs in real-time, all the localization algorithms including anomaly detection will experience a significant performance boost.

Bibliography

[Ahmed et al., 2007] Ahmed, T., Oreshkin, B., and Coates, M. (2007). Machine learning approaches to network anomaly detection. In *Proceedings of the 2nd USENIX workshop on Tackling computer systems problems with machine learning techniques*, pages 1–6. USENIX Association.

[Amer et al., 2013] Amer, M., Goldstein, M., and Abdennadher, S. (2013). Enhancing one-class support vector machines for unsupervised anomaly detection. In *Outlier Detection and Description, Workshop on*.

[Anousaki and Kyriakopoulos, 2004] Anousaki, G. and Kyriakopoulos, K. J. (2004). A dead-reckoning scheme for skid-steered vehicles in outdoor environments. In *Robotics and Automation, IEEE International Conference on*, volume 1, pages 580–585. IEEE.

[Bachrach et al., 2011] Bachrach, A., Prentice, S., He, R., and Roy, N. (2011). Range - robust autonomous navigation in GPS-denied environments. *Journal of Field Robotics*, 28(5):644–666.

[Bar-Shalom et al., 2001] Bar-Shalom, Y., Li, X. R., and Kirubarajan, T. (2001). *Estimation with Applications To Tracking and Navigation*. Wiley and Sons.

[Bishop and Nasrabadi, 2006] Bishop, C. M. and Nasrabadi, N. M. (2006). *Pattern recognition and machine learning*, volume 1. Springer New York.

[Bloesch et al., 2013] Bloesch, M., Gehring, C., Fankhauser, P., Hutter, M., Hoepflinger, M., Siegwart, R., et al. (2013). State estimation for legged robots on unstable and slippery terrain. In *Intelligent Robots and Systems, IEEE/RSJ International Conference on*, pages 6058–6064. IEEE.

[Bodesheim et al., 2013] Bodesheim, P., Freytag, A., Rodner, E., and Denzler, J. (2013). Approximations of gaussian process uncertainties for visual recognition problems. In *Image Analysis*, pages 182–194. Springer.

[Borenstein and Ojeda, 2009] Borenstein, J. and Ojeda, L. (2009). Heuristic reduction of gyro drift in gyro-based vehicle tracking. In *SPIE Defense, Security, and Sensing*, pages 730507–730518.

[Briscoe and Feldman, 2011] Briscoe, E. and Feldman, J. (2011). Conceptual complexity and the bias/variance tradeoff. *Cognition*, 118(1):2–16.

[Brooks and Iagnemma, 2005] Brooks, C. A. and Iagnemma, K. (2005). Vibration-based terrain classification for planetary exploration rovers. *Robotics, IEEE Transactions on*, 21(6):1185–1191.

[Brunner et al., 2013] Brunner, C., Peynot, T., Vidal-Calleja, T., and Underwood, J. (2013). Selective combination of visual and thermal imaging for resilient localization in adverse conditions: Day and night, smoke and fire. *Journal of Field Robotics*, 30(4):641–666.

[Bry et al., 2012] Bry, A., Bachrach, A., and Roy, N. (2012). State estimation for aggressive flight in gps-denied environments using onboard sensing. In *Robotics and Automation, IEEE International Conference on*, pages 1–8.

[Caron et al., 2006] Caron, F., Duflos, E., Pomorski, D., and Vanheeghe, P. (2006). Gps/imu data fusion using multisensor kalman filtering: Introduction of contextual aspects. *Information Fusion*, 7(2):221–230.

[Chalupka et al., 2013] Chalupka, K., Williams, C. K., and Murray, I. (2013). A framework for evaluating approximation methods for gaussian process regression. *The Journal of Machine Learning Research*, 14(1):333–350.

[Chandola et al., 2009] Chandola, V., Banerjee, A., and Kumar, V. (2009). Anomaly detection: A survey. *ACM Computing Surveys*, 41(3):15.

[Chen, 2003] Chen, Z. (2003). Bayesian filtering: From kalman filters to particle filters, and beyond. *Statistics*, 182(1):1–69.

[Chowdhary et al., 2013] Chowdhary, G., Johnson, E. N., Magree, D., Wu, A., and 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.

[Christensen et al., 2008] Christensen, A. L., O'Grady, R., Birattari, M., and Dorigo, M. (2008). Fault detection in autonomous robots based on fault injection and learning. *Autonomous Robots*, 24(1):49–67.

[Chugo et al., 2008] Chugo, D., Kawabata, K., Kaetsu, H., Jia, S., Asama, H., Mishima, T., and Takase, K. (2008). 3d odometry based on body configuration. In *SICE Annual Conference*, pages 695–700.

[Collins and Coyle, 2008] Collins, E. G. and Coyle, E. J. (2008). Vibration-based terrain classification using surface profile input frequency responses. In *Robotics and Automation, IEEE International Conference on*, pages 3276–3283.

[Cook, 2011] Cook, G. (2011). *Mobile robots: Navigation, Control and Remote Sensing*. John Wiley & Sons, Inc.

[Coyle et al., 2011] Coyle, E., Collins, E. G., and Roberts, R. G. (2011). Speed independent terrain classification using singular value decomposition interpolation. In *Robotics and Automation, IEEE International Conference on*, pages 4014–4019.

[Csató and Opper, 2001] Csató, L. and Opper, M. (2001). Sparse representation for gaussian process models. *Advances in neural information processing systems*, pages 444–450.

[Deisenroth, 2010] Deisenroth, M. P. (2010). *Efficient reinforcement learning using Gaussian processes*. PhD thesis, Karlsruhe Institute of Technology.

[Deisenroth and Ng, 2015] Deisenroth, M. P. and Ng, J. W. (2015). Distributed gaussian processes. *arXiv preprint arXiv:1502.02843*.

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

[Do, 2007] Do, C. B. (2007). Gaussian processes. Technical report, Stanford Engineering Everywhere. lecture notes.

[Dua and Du, 2011] Dua, S. and Du, X. (2011). *Data mining and machine learning in cybersecurity*. Taylor & Francis.

[DuPont et al., 2008] DuPont, E. M., Moore, C. A., and Roberts, R. G. (2008). Terrain classification for mobile robots traveling at various speeds: An eigenspace manifold approach. In *Robotics and Automation, IEEE International Conference on*, pages 3284–3289.

[Durrant-Whyte, 2001] Durrant-Whyte, H. F. (2001). Introduction to estimation and the kalman filter. Technical report, Australian Centre for Field Robotics.

[Duvenaud, 2014] Duvenaud, D. (2014). *Automatic model construction with Gaussian processes*. PhD thesis, University of Cambridge.

[Duvenaud et al., 2013] Duvenaud, D., Lloyd, J. R., Grosse, R., Tenenbaum, J. B., and Ghahramani, Z. (2013). Structure discovery in nonparametric regression through compositional kernel search. *arXiv:1302.4922*.

[Ellekilde et al., 2007] Ellekilde, L.-P., Huang, S., Miro, J. V., and Dissanayake, G. (2007). Dense 3d map construction for indoor search and rescue. *Journal of Field Robotics*, 24(1-2):71–89.

[Endo et al., 2007] Endo, D., Okada, Y., Nagatani, K., and Yoshida, K. (2007). Path following control for tracked vehicles based on slip-compensating odometry. In *Intelligent Robots and Systems, IEEE/RSJ International Conference on*, pages 2871–2876.

[Fawcett, 2004] Fawcett, T. (2004). Roc graphs: Notes and practical considerations for researchers. *Machine learning*, 31:1–38.

[Fox et al., 1999] Fox, D., Burgard, W., and Thrun, S. (1999). Markov localization for mobile robots in dynamic environments. *Journal of Artificial Intelligence Research*, pages 391–427.

[Furgale et al., 2013] Furgale, P. T., Newman, P., Triebel, R., and et. al., G. H. (2013). Toward automated driving in cities using close-to-market sensors, an overview of the v-charge project. In *IEEE Intelligent Vehicles Symposium*. To appear.

[Galben, 2011] 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.

[Gao and Li, 2011] Gao, Y. and Li, Y. (2011). Improving gaussian process classification with outlier detection, with applications in image classification. In *Computer Vision–ACCV*, pages 153–164. Springer.

[Gertler, 1998] Gertler, J. (1998). *Fault detection and diagnosis in engineering systems*. CRC press.

[Giguere and Dudek, 2011] Giguere, P. and Dudek, G. (2011). A simple tactile probe for surface identification by mobile robots. *Robotics, IEEE Transactions on*, 27(3):534–544.

[Goel et al., 2000] Goel, P., Dedeoglu, G., Roumeliotis, S. I., and Sukhatme, G. (2000). Fault detection and identification in a mobile robot using multiple model estimation and neural network. In *Robotics and Automation, IEEE International Conference on*, volume 3, pages 2302–2309. IEEE.

[Görner and Stelzer, 2013] Görner, M. and Stelzer, A. (2013). A leg proprioception based 6 dof odometry for statically stable walking robots. *Autonomous Robots*, 34(4):311–326.

[Grewal and Andrews, 2010] Grewal, M. S. and Andrews, A. P. (2010). Applications of kalman filtering in aerospace 1960 to the present [historical perspectives]. *Control Systems, IEEE*, 30(3):69–78.

[Grewal et al., 2007] Grewal, M. S., Weill, L. R., and Andrews, A. P. (2007). *Global positioning systems, inertial navigation, and integration*. John Wiley & Sons.

[Gustafsson et al., 2002] Gustafsson, F., Gunnarsson, F., Bergman, N., Forssell, U., Jansson, J., Karlsson, R., and Nordlund, P.-J. (2002). Particle filters for positioning, navigation, and tracking. *Signal Processing, IEEE Transactions on*, 50(2):425–437.

[Halatci et al., 2007] Halatci, I., Brooks, C. A., and Iagnemma, K. (2007). Terrain classification and classifier fusion for planetary exploration rovers. In *Aerospace, IEEE International Conference on*, pages 1–11.

[Hall and Llinas, 1997] Hall, D. L. and Llinas, J. (1997). An introduction to multisensor data fusion. *Proceedings of the IEEE*, 85(1):6–23.

[Hoffmann et al., 2014] Hoffmann, M., Štěpánová, K., and Reinstein, M. (2014). The effect of motor action and different sensory modalities on terrain classification in a quadruped robot running with multiple gaits. *Robotics and Autonomous Systems*, 62(12):1790–1798.

[Howard and Seraji, 2000] Howard, A. and Seraji, H. (2000). Real-time assessment of terrain traversability for autonomous rover navigation. In *Intelligent Robots and Systems, IEEE/RSJ International Conference on*, volume 1, pages 58–63.

[Hwang et al., 2010] Hwang, I., Kim, S., Kim, Y., and Seah, C. E. (2010). A survey of fault detection, isolation, and reconfiguration methods. *Control Systems Technology, IEEE Transactions on*, 18(3):636–653.

[Jeni et al., 2013] Jeni, L., Cohn, J. F., De La Torre, F., et al. (2013). Facing imbalanced data–recommendations for the use of performance metrics. In *Affective Computing and Intelligent Interaction, IEEE Humaine Association Conference on*, pages 245–251.

[Jitpakdee and Maneewarn, 2008] Jitpakdee, R. and Maneewarn, T. (2008). Neural networks terrain classification using inertial measurement unit for an autonomous vehicle. In *SICE Annual Conference*, pages 554–558.

[Julier and Uhlmann, 1997] Julier, S. J. and Uhlmann, J. K. (1997). New extension of the kalman filter to nonlinear systems. In *AeroSense*, pages 182–193. International Society for Optics and Photonics.

[Jwo and Cho, 2010] Jwo, D.-J. and Cho, T.-S. (2010). Critical remarks on the linearised and extended kalman filters with geodetic navigation examples. *Measurement*, 43(9):1077–1089.

[Kalman, 1960] Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. *Journal of Fluids Engineering*, 82(1):35–45.

[Kemmler et al., 2013] Kemmler, M., Rodner, E., Wacker, E.-S., and Denzler, J. (2013). One-class classification with gaussian processes. *Pattern Recognition*, 46(12):3507–3518.

[Kleiner and Dornhege, 2007] Kleiner, A. and Dornhege, C. (2007). Real-time localization and elevation mapping within urban search and rescue scenarios: Field reports. *Journal of Field Robotics*, 24(8-9):723–745.

[Knorn and Leith, 2008] Knorn, F. and Leith, D. J. (2008). Adaptive kalman filtering for anomaly detection in software appliances. In *INFOCOM Workshops, IEEE*, pages 1–6. IEEE.

[Komma et al., 2009] Komma, P., Weiss, C., and Zell, A. (2009). Adaptive bayesian filtering for vibration-based terrain classification. In *Robotics and Automation, IEEE International Conference on*, pages 3307–3313.

[Konolige et al., 2011] Konolige, K., Agrawal, M., and Sola, J. (2011). Large-scale visual odometry for rough terrain. In *Robotics Research*, volume 66, pages 201–212. Springer Berlin Heidelberg.

[Kowalska and Peel, 2012] Kowalska, K. and Peel, L. (2012). Maritime anomaly detection using gaussian process active learning. In *Information Fusion, International Conference on*, pages 1164–1171. IEEE.

[Kruijff et al., 2012] Kruijff, G.-J. M., Pirri, F., Gianni, M., Papadakis, P., Pizzoli, M., Sinha, A., Pianese, E., Corrao, S., Priori, F., Febrini, S., Angeletti, S., Tretyakov, V., and Linder, T. (2012). Rescue robots at earthquake-hit mirandola, italy: a field report. In *Safety Security and Rescue Robotics, IEEE International Symposium of*.

[Kubelka et al., 2015] Kubelka, V., Oswald, L., Pomerleau, F., Colas, F., Svoboda, T., and Reinstein, M. (2015). Robust data fusion of multimodal sensory information for mobile robots. *Journal of Field Robotics*, 32(4):447–473.

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

[Kümmerle et al., 2013] Kümmerle, R., Ruhnke, M., Steder, B., Stachniss, C., and Burgard, W. (2013). A navigation system for robots operating in crowded urban environments. In *Robotics and Automation, IEEE International Conference on*.

[Kuss and Rasmussen, 2005] Kuss, M. and Rasmussen, C. E. (2005). Assessing approximate inference for binary gaussian process classification. *The Journal of Machine Learning Research*, 6:1679–1704.

[Lamon and Siegwart, 2004] Lamon, P. and Siegwart, R. (2004). Inertial and 3d-odometry fusion in rough terrain - towards real 3d navigation. In *Intelligent Robots and Systems, IEEE/RSJ International Conference on*, volume 2, pages 1716–1721.

[Laviola, 2003] Laviola, J. J. (2003). A comparison of unscented and extended kalman filtering for estimating quaternion motion. In *American Control Conference*, volume 3, pages 2435–2440. IEEE.

[Laxhammar et al., 2009] Laxhammar, R., Falkman, G., and Sviestins, E. (2009). Anomaly detection in sea traffic-a comparison of the gaussian mixture model and the kernel density estimator. In *Information Fusion, International Conference on*, pages 756–763. IEEE.

[Li and Mourikis, 2013] Li, M. and Mourikis, A. I. (2013). High-precision, consistent EKF-based visual–inertial odometry. *The International Journal of Robotics Research*, 32(6):690–711.

[Ma et al., 2015] Ma, J., Bajracharya, M., Susca, S., Matthies, L., and Malchano, M. (2015). Real-time pose estimation of a dynamic quadruped in GPS-denied environments for 24-hour operation. *The International Journal of Robotics Research*, pages 1–23.

[Matheron, 1963] Matheron, G. (1963). Principles of geostatistics. *Economic geology*, 58(8):1246–1266.

[Maybeck, 1982] Maybeck, P. S. (1982). *Stochastic models, estimation, and control*, volume 3. Academic press.

[McDonald et al., 2013] McDonald, J., Kaess, M., Cadena, C., Neira, J., and Leonard, J. J. (2013). Real-time 6-dof multi-session visual slam over large-scale environments. *Robotics and Autonomous Systems*, 61(10):1144–1158.

[Mendoza et al., 2012] Mendoza, J. P., Veloso, M., and Simmons, R. (2012). Mobile robot fault detection based on redundant information statistics. In *Safety in Human–Robot Coexistence and Interaction, Workshop on*. Citeseer.

[Michael et al., 2012] Michael, N., Shen, S., Mohta, K., Mulgaonkar, Y., Kumar, V., Nagatani, K., Okada, Y., Kiribayashi, S., Otake, K., Yoshida, K., Ohno, K., Takeuchi, E., and Tadokoro, S. (2012). Collaborative mapping of an earthquake-damaged building via ground and aerial robots. *Journal of Field Robotics*, 29(5):832–841.

[Milella et al., 2015] Milella, A., Reina, G., and Underwood, J. (2015). A self-learning framework for statistical ground classification using radar and monocular vision. *Journal of Field Robotics*, 32(1):20–41.

[Minka, 2001] Minka, T. P. (2001). *A family of algorithms for approximate Bayesian inference*. PhD thesis, Massachusetts Institute of Technology.

[Morales et al., 2008] Morales, Y., Takeuchi, E., and Tsubouchi, T. (2008). Vehicle localization in outdoor woodland environments with sensor fault detection. In *Robotics and Automation, IEEE International Conference on*, pages 449–454. IEEE.

[Morales et al., 2009] Morales, Y., Tsubouchi, T., and Yuta, S. (2009). Vehicle 3d localization in mountainous woodland environments. In *Intelligent Robots and Systems, IEEE/RSJ International Conference on*, pages 3588–3594.

[Munguia and Grau, 2014] Munguia, R. and Grau, A. (2014). A practical method for implementing an attitude and heading reference system. *International Journal of Advanced Robotic Systems*, 11(62).

[Murray, 2008] Murray, I. (2008). Introduction to gaussian processes. CSC2515 lecture notes, Fall 2008, Dept. Computer Science, University of Toronto.

[Nagatani et al., 2011] Nagatani, K., Okada, Y., Tokunaga, N., Kiribayashi, S., Yoshida, K., Ohno, K., Takeuchi, E., Tadokoro, S., Akiyama, H., Noda, I., Yoshida, T., and 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.

[Ndong and Salamatian, 2011] Ndong, J. and Salamatian, K. (2011). A robust anomaly detection technique using combined statistical methods. In *Communication Networks and Services Research, Conference on*, pages 101–108. IEEE.

[Nowozin, 2014] Nowozin, S. (2014). Optimal decisions from probabilistic models: the intersection-over-union case. In *Computer Vision and Pattern Recognition, IEEE Conference on*, pages 548–555. IEEE.

[Okada et al., 2011] Okada, Y., Nagatani, K., Yoshida, K., Tadokoro, S., Yoshida, T., and Koyanagi, E. (2011). Shared autonomy system for tracked vehicles on rough terrain based on continuous three-dimensional terrain scanning. *Journal of Field Robotics*, 28(6):875–893.

[Paciorek and Schervish, 2004] Paciorek, C. and Schervish, M. (2004). Nonstationary covariance functions for gaussian process regression. *Advances in neural information processing systems*, 16:273–280.

[Pettersson, 2005] Pettersson, O. (2005). Execution monitoring in robotics: A survey. *Robotics and Autonomous Systems*, 53(2):73–88.

[Quiñonero-Candela and Rasmussen, 2005] Quiñonero-Candela, J. and Rasmussen, C. E. (2005). A unifying view of sparse approximate gaussian process regression. *The Journal of Machine Learning Research*, 6:1939–1959.

[Ramirez-Padron et al., 2013] Ramirez-Padron, R., Mederos, B., and Gonzalez, A. J. (2013). Novelty detection using sparse online gaussian processes for visual object recognition. In *FLAIRS, International Conference*.

[Rasmussen, 2004] Rasmussen, C. E. (2004). Gaussian processes in machine learning. In *Advanced lectures on machine learning*, pages 63–71. Springer.

[Rasmussen, 2006] Rasmussen, C. E. (2006). Covariance functions and classification. Gaussian Processes in Practice lecture notes, July 2006, Bletchley Park.

[Rasmussen and Williams, 2006] Rasmussen, C. E. and Williams, C. K. (2006). *Gaussian processes for machine learning*. the MIT Press.

[Reece et al., 2009] Reece, S., Garnett, R., Osborne, M., and Roberts, S. (2009). Anomaly detection and removal using nonstationary gaussian processes. *University of Oxford, Technical Report*.

[Reinstein and Hoffmann, 2011] Reinstein, M. and Hoffmann, M. (2011). Dead reckoning in a dynamic quadruped robot: Inertial navigation system aided by a legged odometer. In *Robotics and Automation, IEEE International Conference on*, pages 617–624. IEEE.

[Reinstein and Hoffmann, 2013] Reinstein, M. and Hoffmann, M. (2013). Dead reckoning in a dynamic quadruped robot based on multimodal proprioceptive sensory information. *Robotics, IEEE Transactions on*, 29(2):563–571.

[Reinstein et al., 2013] Reinstein, M., Kubelka, V., and Zimmermann, K. (2013). Terrain adaptive odometer for mobile skid-steer robots. In *Robotics and Automation, IEEE International Conference on*, pages 4691–4696.

[Rodriguez F et al., 2009] Rodriguez F, S. A., Fremont, V., and Bonnifait, P. (2009). An experiment of a 3d real-time robust visual odometry for intelligent vehicles. In *Intelligent Transportation Systems, International Conference on*, pages 1–6.

[Roumeliotis et al., 1998] Roumeliotis, S. I., Sukhatme, G., and Bekey, G. A. (1998). Sensor fault detection and identification in a mobile robot. In *Intelligent Robots and Systems, IEEE/RSJ International Conference on*, volume 3, pages 1383–1388. IEEE.

[Roumeliotis et al., 1999] Roumeliotis, S. I., Sukhatme, G. S., and Bekey, G. A. (1999). Circumventing dynamic modeling: Evaluation of the error-state kalman filter applied to mobile robot localization. In *Robotics and Automation, IEEE International Conference on*, volume 2, pages 1656–1663. IEEE.

[Sagha et al., 2013] Sagha, H., Bayati, H., Millan, J. d. R., and Chavarriaga, R. (2013). On-line anomaly detection and resilience in classifier ensembles. *Pattern Recognition Letters*.

[Sakai et al., 2009] Sakai, A., Tamura, Y., and Kuroda, Y. (2009). An efficient solution to 6DOF localization using unscented kalman filter for planetary rovers. In *Intelligent Robots and Systems, IEEE/RSJ International Conference on*, pages 4154–4159.

[Santamaria-Navarro et al., 2015] Santamaria-Navarro, À., Teniente, E. H., Morta, M., and Andrade-Cetto, J. (2015). Terrain classification in complex three-dimensional outdoor environments. *Journal of Field Robotics*, 32(1):42–60.

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

[Serradilla et al., 2011] Serradilla, J., Shi, J., and Morris, A. (2011). Fault detection based on gaussian process latent variable models. *Chemometrics and Intelligent Laboratory Systems*, 109(1):9–21.

[Shen et al., 2011] Shen, J., Tick, D., and Gans, N. (2011). Localization through fusion of discrete and continuous epipolar geometry with wheel and imu odometry. In *American Control Conference*, pages 1292–1298.

[Shin, 2005] Shin, E. H. (2005). *Estimation Techniques for Low-Cost Inertial Navigation*. PhD thesis, University of Calgary, Calgary.

[Siciliano and Khatib, 2008] Siciliano, B. and Khatib, O. (2008). *Springer handbook of robotics*. Springer Science & Business Media.

[Simanek et al., 2015] Simanek, J., Reinstein, M., and Kubelka, V. (2015). Evaluation of the EKF-based estimation architectures for data fusion in mobile robots. *Mechatronics, IEEE/ASME Transactions on*, 20(2):985–990.

[Smith et al., 2014] Smith, M., Reece, S., Roberts, S., Psorakis, I., and Rezek, I. (2014). Maritime abnormality detection using gaussian processes. *Knowledge and information systems*, 38(3):717–741.

[Soule et al., 2005] Soule, A., Salamatian, K., and Taft, N. (2005). Combining filtering and statistical methods for anomaly detection. In *Internet Measurement, Conference on*, pages 31–31. USENIX Association.

[Sundvall and Jensfelt, 2006] Sundvall, P. and Jensfelt, P. (2006). Fault detection for mobile robots using redundant positioning systems. In *Robotics and Automation, IEEE International Conference on*, pages 3781–3786. IEEE.

[Suzuki et al., 2010] Suzuki, T., Kitamura, M., Amano, Y., and Hashizume, T. (2010). 6-dof localization for a mobile robot using outdoor 3d voxel maps. In *Intelligent Robots and Systems, IEEE/RSJ International Conference on*, pages 5737–5743.

[Suzuki et al., 2011] Suzuki, T., Kitamura, M., Amano, Y., and Hashizume, T. (2011). High-accuracy gps and glonass positioning by multipath mitigation using omnidirectional infrared camera. In *Robotics and Automation, IEEE International Conference on*, pages 311–316. IEEE.

[Thrun et al., 2005] Thrun, S., Burgard, W., Fox, D., et al. (2005). *Probabilistic robotics*, volume 1. MIT press Cambridge.

[Tick et al., 2012] Tick, D., Rahman, T., Busso, C., and Gans, N. (2012). Indoor robotic terrain classification via angular velocity based hierarchical classifier selection. In *Robotics and Automation, IEEE International Conference on*, pages 3594–3600.

[Ting et al., 2007] Ting, J.-A., Theodorou, E., and Schaal, S. (2007). A kalman filter for robust outlier detection. In *Intelligent Robots and Systems, IEEE/RSJ International Conference on*, pages 1514–1519. IEEE.

[Vandyke et al., 2004] Vandyke, M. C., Schwartz, J. L., and Hall, C. D. (2004). Unscented kalman filtering for spacecraft attitude state and parameter estimation. In *Space Flight Mechanics, AAS/AIAA Conference on*.

[Weiss et al., 2007] Weiss, C., Fechner, N., Stark, M., and Zell, A. (2007). Comparison of different approaches to vibration-based terrain classification. Technical report.

[Weiss et al., 2006] Weiss, C., Frohlich, H., and Zell, A. (2006). Vibration-based terrain classification using support vector machines. In *Intelligent Robots and Systems, IEEE/RSJ International Conference on*, pages 4429–4434.

[Weiss et al., 2008] Weiss, C., Tamimi, H., and Zell, A. (2008). A combination of vision- and vibration-based terrain classification. In *Intelligent Robots and Systems, IEEE/RSJ International Conference on*, pages 2204–2209.

[Williams and Barber, 1998] Williams, C. K. and Barber, D. (1998). Bayesian classification with gaussian processes. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 20(12):1342–1351.

[Yi et al., 2007] Yi, J., Zhang, J., Song, D., and Jayasuriya, S. (2007). Imu-based localization and slip estimation for skid-steered mobile robots. In *Intelligent Robots and Systems, IEEE/RSJ International Conference on*, pages 2845–2850.

[Yoshida et al., 2010] Yoshida, T., Irie, K., Koyanagi, E., and Tomono, M. (2010). A sensor platform for outdoor navigation using gyro-assisted odometry and roundly-swinging 3d laser scanner. In *Intelligent Robots and Systems, IEEE/RSJ International Conference on*, pages 1414–1420.

[Zhuang et al., 2011] Zhuang, Y., Yang, S., Li, X., and Wang, W. (2011). 3d-laser-based visual odometry for autonomous mobile robot in outdoor environments. In *Awareness Science and Technology, International Conference on*, pages 133–138. IEEE.

[Zimmermann et al., 2015] Zimmermann, K., Zuzánek, P., Reinstein, M., Petrıcek, T., and Hlavác, V. (2015). Adaptive traversability of partially occluded obstacles. In *Robotics and Automation, IEEE International Conference on*.

Publications of the Author

Thesis related publications

Articles in peer-reviewed journals

[A.1] Simanek, J., Reinstein, M., Kubelka. V. (2015). *Evaluation of the EKF-based Estimation Architectures for Data Fusion in Mobile Robots.* IEEE-ASME Transactions on Mechatronics.

[A.2] Simanek, J., Kubelka, V., Reinstein, M. (2015). *Improving multi-modal data fusion by anomaly detection.* Autonomous Robots.

Conference proceedings

[A.3] Alam, M., Sipos, M., Rohac, J., Simanek, J. (2015). *Calibration of a multi-sensor inertial measurement unit with modified sensor frame.* Industrial Technology, IEEE International Conference on, Spain.

[A.4] Sipos, M., Simanek, J., Novacek, P., Popelka, J., Rohac, J. (2013). *Modular Navigation System for Unmanned Aerial Vehicles.* Military Technologies, International Conference on, Czech Republic.

[A.5] Rohac, J., Sipos, M., Simanek, J., Teren, O. (2012). *Inertial Reference Unit in a Directional Gyro Mode of Operation.* IEEE Sensors, International Conference on, Taiwan. (co-authorship 30%).

[A.6] Novacek, P., Sipos, M., Popelka, J., Simanek, J., Rohac, J. (2012). *Modular Navigation System for Unmanned Aerial Vehicles - Flight Data Monitoring and Recording.* Diagnostic and Reliability of Onboard Aircraft System, Czech Republic.

[A.7] Simanek, J. (2012). *Analyses of suboptimal models for INS/GPS navigation algorithms.* The 8th AIAA-Pegasus Student Conference, ENSMA, Poitiers, France.

Other publications

Articles in peer-reviewed journals

[A.8] Novacek, P., Rohac, J., Simanek, J., Ripka, P. (2013). *Metal Detector Signal Imprints of Detected Objects.* IEEE Transactions on Magnetics. (co-authorship 5%).

Accepted

[A.9] Rohac, J., Sipos, M., Simanek, J. (2015). *Calibration of the Low-cost Triaxial Inertial Sensors.* IEEE Instrumentation & Measurement Magazine.

Submitted and under review

[A.10] Hoffmann, M., Simanek, J. (2014). *Passive compliant joints in legged locomotion: faster learning, superior energy efficiency, and enhanced sensing.* Robotics and Autonomous Systems.

Responses to author's publications

Simanek, J., Kubelka, V., Reinstein, M. (2015). *Improving multi-modal data fusion by anomaly detection.* Autonomous Robots.

- Chen, Z., Zhang, F., Qu, X., Liang, B. (2015). *Fast Measurement and Reconstruction of Large Workpieces with Freeform Surfaces by Combining Local Scanning and Global Position Data.* Sensors.

Simanek, J., Reinstein, M., Kubelka, V. (2015). *Evaluation of the EKF-based Estimation Architectures for Data Fusion in Mobile Robots.* IEEE-ASME Transactions on Mechatronics.

- Benziane, L. (2015). *Attitude estimation & control of autonomous aerial vehicles.* PhD thesis, Universite de Versailles-Saint Quentin en Yvelines, Versailles.
- Chen, S. L., Kamaldin, N., Teo, T., Liang, W., Teo, C., Yang, G., Tan, K. (2015). *Towards Comprehensive Modeling and Large Angle Tracking Control of a Limited Angle Torque Actuator with Cylindrical Halbach.* Mechatronics, IEEE-ASME Transactions on.

- Atia, M., Liu, S., Nematallah, H., Karamat, T. B., Noureldin, A. (2015). *Integrated Indoor Navigation System for Ground Vehicles With Automatic 3-D Alignment and Position Initialization.* Vehicular Technology, IEEE Transactions on.
- Zuzanek, P., Zimmermann, K., Hlavac, V. (2014). *Accepted Autonomy for Search and Rescue Robotics.* Modelling and Simulation for Autonomous Systems: First International Workshop, MESAS, Springer.

Rohac, J., Sipos, M., Simanek, J., Teren, O. (2012). *Inertial Reference Unit in a Directional Gyro Mode of Operation.* IEEE Sensors, International Conference on, Taiwan.

- Vaispacher, T., Breda, R., Madarasz, L. (2015). *Integration architectures of navigation systems for unmanned vehicles.* Applied Machine Intelligence and Informatics, IEEE International Symposium on.

Simanek, J. (2012). *Analyses of suboptimal models for INS/GPS navigation algorithms.* The 8th AIAA-Pegasus Student Conference, ENSMA, Poitiers, France.

- Abhijit, S., et al. (2014). *A Precise State Transition Model for Aircraft Navigation.* Geomatica.

Appendix A

A.1 Gaussian processes classification results

In the following, detailed results from the performance analysis described in Section 6.3.1 are shown.

Feat. | Negatives	#	TP	FP	TN	FN	PR	G	AUC
norm | balanced	1	163	159	7585	6	0.51	0.9720	0.9839
innov | balanced	2	161	78	7666	8	0.67	0.9711	0.9818
norm | 0.1	3	161	88	7656	8	0.65	0.9705	0.9839
innov | 0.1	4	158	54	7690	11	0.75	0.9635	0.9823
norm | 0.3	5	161	78	7666	8	0.67	0.9711	0.9839
innov | 0.3	6	161	88	7656	8	0.65	0.9705	0.9822
norm | 0.5	7	161	78	7666	8	0.67	0.9711	0.9839
innov | 0.5	8	155	40	7704	14	0.79	0.9552	0.9823
norm | 1	9	161	78	7666	8	0.67	0.9711	0.9839
innov | 1	10	161	75	7669	8	0.68	0.9713	0.9817

Table A.1: Evaluation of sparse subsampled implementations - VO attitude. Row caption specify the following parameter settings: Features (Feat.) {norm of innovations, innovation vector}, number of negative training samples (Negatives) {balacned or ratio of all negative training samples}

Feat. | Negatives	#	TP	FP	TN	FN	PR	G	AUC
norm | balanced	1	25	78	1499	2	0.24	0.9382	0.9794
innov | balanced	2	17	133	1444	10	0.11	0.7593	0.7783
norm | 0.1	3	25	77	1500	2	0.25	0.9385	0.9794
innov | 0.1	4	19	132	1445	8	0.13	0.8030	0.8079
norm | 0.3	5	24	70	1507	3	0.26	0.9216	0.9793
innov | 0.3	6	14	91	1486	13	0.13	0.6990	0.8240
norm | 0.5	7	25	77	1500	2	0.25	0.9385	0.9793
innov | 0.5	8	15	93	1484	12	0.14	0.7230	0.8283
norm | 1	9	24	72	1505	3	0.25	0.9210	0.9794
innov | 1	10	14	66	1511	13	0.17	0.7049	0.8446

Table A.2: Evaluation of sparse subsampled implementations - Laser attitude. Row caption specify the following parameter settings: Features (Feat.) {norm of innovations, innovation vector}, number of negative training samples (Negatives) {balacned or ratio of all negative training samples}

Feat. | Negatives	#	TP	FP	TN	FN	PR	G	AUC
norm | balanced	1	8	128	1468	0	0.06	0.9591	0.9929
innov | balanced	2	8	160	1436	0	0.05	0.9486	0.9793
norm | 0.1	3	8	94	1502	0	0.08	0.9701	0.9921
innov | 0.1	4	8	69	1527	0	0.10	0.9781	0.9879
norm | 0.3	5	8	96	1500	0	0.08	0.9695	0.9919
innov | 0.3	6	8	74	1522	0	0.10	0.9765	0.9922
norm | 0.5	7	8	91	1505	0	0.08	0.9711	0.9919
innov | 0.5	8	8	75	1521	0	0.10	0.9762	0.9928
norm | 1	9	8	106	1490	0	0.07	0.9662	0.9920
innov | 1	10	8	77	1519	0	0.09	0.9756	0.9931

Table A.3: Evaluation of sparse subsampled implementations - Laser position. Row caption specify the following parameter settings: Features (Feat.) {norm of innovations, innovation vector}, number of negative training samples (Negatives) {balacned or ratio of all negative training samples}

Feat. | #Points | Range | Distribution	#	TP	FP	TN	FN	PR	G	AUC
norm | 5 | min-max | rand.	1	45	1	7743	124	0.98	0.5160	0.2913
innov | 5 | min-max | rand.	2	148	250	7494	21	0.37	0.9206	0.9324
norm | 10 | min-max | rand.	3	160	39	7705	9	0.80	0.9706	0.9839
innov | 10 | min-max | rand.	4	161	73	7671	8	0.69	0.9714	0.9828
norm | 15 | min-max | rand.	5	159	38	7706	10	0.81	0.9676	0.9839
innov | 15 | min-max | rand.	6	161	75	7669	8	0.68	0.9713	0.9817
norm | 5 | ±2 | rand.	7	161	77	7667	8	0.68	0.9712	0.9839
innov | 5 | ±2 | rand.	8	151	465	7279	18	0.25	0.9164	0.9443
norm | 10 | ±2 | rand.	9	159	39	7705	10	0.80	0.9675	0.9839
innov | 10 | ±2 | rand.	10	162	227	7517	7	0.42	0.9646	0.9840
norm | 15 | ±2 | rand.	11	161	77	7667	8	0.68	0.9712	0.9839
innov | 15 | ±2 | rand.	12	161	72	7672	8	0.69	0.9715	0.9816
norm | 5 | ±10 | rand.	13	160	44	7700	9	0.78	0.9702	0.9840
innov | 5 | ±10 | rand.	14	93	7664	80	76	0.01	0.0754	0.4741
norm | 10 | ±10 | rand.	15	160	44	7700	9	0.78	0.9702	0.9840
innov | 10 | ±10 | rand.	16	0	0	7744	169	-	0.0000	0.5835
norm | 15 | ±10 | rand.	17	160	44	7700	9	0.78	0.9702	0.9840
innov | 15 | ±10 | rand.	18	28	1	7743	141	0.97	0.4070	0.6476
norm | 5 | min-max | equi.	19	160	39	7705	9	0.80	0.9706	0.9839
innov | 5 | min-max | equi.	20	90	841	6903	79	0.10	0.6890	0.5722
norm | 10 | min-max | equi.	21	159	39	7705	10	0.80	0.9675	0.9839
innov | 10 | min-max | equi.	22	160	51	7693	9	0.76	0.9698	0.9820
norm | 15 | min-max | equi.	23	161	78	7666	8	0.67	0.9711	0.9839
innov | 15 | min-max | equi.	24	161	75	7669	8	0.68	0.9713	0.9817
norm | 5 | ±2 | equi.	25	161	103	7641	8	0.61	0.9695	0.9840
innov | 5 | ±2 | equi.	26	161	103	7641	8	0.61	0.9695	0.9839
norm | 10 | ±2 | equi.	27	159	37	7707	10	0.81	0.9676	0.9839
innov | 10 | ±2 | equi.	28	134	34	7710	35	0.80	0.8885	0.9780
norm | 15 | ±2 | equi.	29	159	38	7706	10	0.81	0.9676	0.9839
innov | 15 | ±2 | equi.	30	161	74	7670	8	0.69	0.9714	0.9817
norm | 5 | ±10 | equi.	31	161	103	7641	8	0.61	0.9695	0.9840
innov | 5 | ±10 | equi.	32	161	103	7641	8	0.61	0.9695	0.9840
norm | 10 | ±10 | equi.	33	17	0	7744	152	1.00	0.3172	0.9840
innov | 10 | ±10 | equi.	34	0	0	7744	169	-	0.0000	0.4725
norm | 15 | ±10 | equi.	35	161	103	7641	8	0.61	0.9695	0.9840

Table A.4: Evaluation of FITC implementations - VO attitude. Row captions specify the following parameter settings: features (Feat.) {norm of innovations, innovation vector}, number of inducing points (#Points) {distributed in the range}, range of inducting points (Range), distribution of the points (Dist.){uniform random or equidistant}.

Feat. | #Points | Range | Distribution	#	TP	FP	TN	FN	PR	G	AUC
norm | 5 | min-max | rand.	1	24	70	1507	3	0.26	0.9216	0.9794
innov | 5 | min-max | rand.	2	12	156	1421	15	0.07	0.6328	0.6280
norm | 10 | min-max | rand.	3	24	72	1505	3	0.25	0.9210	0.9794
innov | 10 | min-max | rand.	4	14	66	1511	13	0.17	0.7049	0.8444
norm | 15 | min-max | rand.	5	24	72	1505	3	0.25	0.9210	0.9794
innov | 15 | min-max | rand.	6	14	66	1511	13	0.17	0.7049	0.8446
norm | 5 | ±2 | rand.	7	25	72	1505	2	0.26	0.9400	0.9795
innov | 5 | ±2 | rand.	8	16	145	1432	11	0.10	0.7336	0.6499
norm | 10 | ±2 | rand.	9	24	71	1506	3	0.25	0.9213	0.9794
innov | 10 | ±2 | rand.	10	17	91	1486	10	0.16	0.7703	0.8205
norm | 15 | ±2 | rand.	11	24	72	1505	3	0.25	0.9210	0.9794
innov | 15 | ±2 | rand.	12	19	101	1476	8	0.16	0.8116	0.8024
norm | 5 | ±10 | rand.	13	25	78	1499	2	0.24	0.9382	0.9795
innov | 5 | ±10 | rand.	14	27	1577	0	0	0.02	0.0000	0.6291
norm | 10 | ±10 | rand.	15	25	78	1499	2	0.24	0.9382	0.9795
innov | 10 | ±10 | rand.	16	27	1577	0	0	0.02	0.0000	0.3750
norm | 15 | ±10 | rand.	17	25	78	1499	2	0.24	0.9382	0.9795
innov | 15 | ±10 | rand.	18	27	1577	0	0	0.02	0.0000	0.6331
norm | 5 | min-max | equi.	19	25	72	1505	2	0.26	0.9400	0.9794
innov | 5 | min-max | equi.	20	13	252	1325	14	0.05	0.6360	0.6723
norm | 10 | min-max | equi.	21	24	72	1505	3	0.25	0.9210	0.9794
innov | 10 | min-max | equi.	22	14	66	1511	13	0.17	0.7049	0.8440
norm | 15 | min-max | equi.	23	24	72	1505	3	0.25	0.9210	0.9794
innov | 15 | min-max | equi.	24	14	66	1511	13	0.17	0.7049	0.8446
norm | 5 | ±2 | equi.	25	26	102	1475	1	0.20	0.9490	0.9795
innov | 5 | ±2 | equi.	26	26	102	1475	1	0.20	0.9490	0.9796
norm | 10 | ±2 | equi.	27	24	70	1507	3	0.26	0.9216	0.9795
innov | 10 | ±2 | equi.	28	9	78	1499	18	0.10	0.5629	0.5702
norm | 15 | ±2 | equi.	29	25	72	1505	2	0.26	0.9400	0.9794
innov | 15 | ±2 | equi.	30	14	67	1510	13	0.17	0.7046	0.8476
norm | 5 | ±10 | equi.	31	26	102	1475	1	0.20	0.9490	0.9795
innov | 5 | ±10 | equi.	32	26	102	1475	1	0.20	0.9490	0.9795
norm | 10 | ±10 | equi.	33	27	1576	1	0	0.02	0.0252	0.0205
innov | 10 | ±10 | equi.	34	27	1577	0	0	0.02	0.0000	0.5000
norm | 15 | ±10 | equi.	35	26	102	1475	1	0.20	0.9490	0.9795

Table A.5: Evaluation of FITC implementations - Laser attitude. Row captions specify the following parameter settings: features (Feat.) {norm of innovations, innovation vector}, number of inducing points (#Points) {distributed in the range}, range of inducting points (Range), distribution of the points (Dist.){uniform random or equidistant}.

Feat. | #Points | Range | Distribution	#	TP	FP	TN	FN	PR	G	AUC
norm | 5 | min-max | rand.	1	8	98	1498	0	0.08	0.9688	0.9920
innov | 5 | min-max | rand.	2	8	74	1522	0	0.10	0.9765	0.9930
norm | 10 | min-max | rand.	3	8	106	1490	0	0.07	0.9662	0.9920
innov | 10 | min-max | rand.	4	8	77	1519	0	0.09	0.9756	0.9930
norm | 15 | min-max | rand.	5	8	106	1490	0	0.07	0.9662	0.9920
innov | 15 | min-max | rand.	6	8	77	1519	0	0.09	0.9756	0.9931
norm | 5 | ±2 | rand.	7	8	94	1502	0	0.08	0.9701	0.9920
innov | 5 | ±2 | rand.	8	8	194	1402	0	0.04	0.9373	0.9583
norm | 10 | ±2 | rand.	9	8	90	1506	0	0.08	0.9714	0.9917
innov | 10 | ±2 | rand.	10	8	64	1532	0	0.11	0.9797	0.9946
norm | 15 | ±2 | rand.	11	8	103	1493	0	0.07	0.9672	0.9920
innov | 15 | ±2 | rand.	12	8	71	1525	0	0.10	0.9775	0.9920
norm | 5 | ±10 | rand.	13	8	91	1505	0	0.08	0.9711	0.9920
innov | 5 | ±10 | rand.	14	7	1596	0	1	0.00	0.0000	0.8634
norm | 10 | ±10 | rand.	15	8	91	1505	0	0.08	0.9711	0.9920
innov | 10 | ±10 | rand.	16	8	1596	0	0	0.00	0.0000	0.8873
norm | 15 | ±10 | rand.	17	8	89	1507	0	0.08	0.9717	0.9925
innov | 15 | ±10 | rand.	18	8	1596	0	0	0.00	0.0000	0.1437
norm | 5 | min-max | equi.	19	8	89	1507	0	0.08	0.9717	0.9920
innov | 5 | min-max | equi.	20	8	66	1530	0	0.11	0.9791	0.9922
norm | 10 | min-max | equi.	21	8	106	1490	0	0.07	0.9662	0.9920
innov | 10 | min-max | equi.	22	8	77	1519	0	0.09	0.9756	0.9931
norm | 15 | min-max | equi.	23	8	106	1490	0	0.07	0.9662	0.9920
innov | 15 | min-max | equi.	24	8	77	1519	0	0.09	0.9756	0.9931
norm | 5 | ±2 | equi.	25	8	94	1502	0	0.08	0.9701	0.9920
innov | 5 | ±2 | equi.	26	8	94	1502	0	0.08	0.9701	0.9922
norm | 10 | ±2 | equi.	27	8	105	1491	0	0.07	0.9665	0.9920
innov | 10 | ±2 | equi.	28	8	98	1498	0	0.08	0.9688	0.9876
norm | 15 | ±2 | equi.	29	8	88	1508	0	0.08	0.9720	0.9920
innov | 15 | ±2 | equi.	30	8	86	1510	0	0.09	0.9727	0.9933
norm | 5 | ±10 | equi.	31	8	94	1502	0	0.08	0.9701	0.9920
innov | 5 | ±10 | equi.	32	8	94	1502	0	0.08	0.9701	0.9920
norm | 10 | ±10 | equi.	33	7	1591	5	1	0.00	0.0524	0.0080
innov | 10 | ±10 | equi.	34	0	7	1589	8	0.00	0.0000	0.2762
norm | 15 | ±10 | equi.	35	8	94	1502	0	0.08	0.9701	0.9920

Table A.6: Evaluation of FITC implementations - Laser position. Row captions specify the following parameter settings: features (Feat.) {norm of innovations, innovation vector}, number of inducing points (#Points) {distributed in the range}, range of inducting points (Range), distribution of the points (Dist.){uniform random or equidistant}.

Feat. | Covariance function	#	TP	FP	TN	FN	PR	G	AUC
norm | SE	1	163	159	7585	6	0.51	0.9720	0.9839
innov | SE	2	161	78	7666	8	0.67	0.9711	0.9818
norm | RQ	3	163	159	7585	6	0.51	0.9720	0.9839
innov | RQ	4	161	82	7662	8	0.66	0.9709	0.9815
norm | Matern 1/2	5	163	160	7584	6	0.50	0.9719	0.9836
innov | Matern 1/2	6	159	124	7620	10	0.56	0.9622	0.9824
norm | Matern 3/2	7	163	160	7584	6	0.50	0.9719	0.9838
innov | Matern 3/2	8	162	148	7596	7	0.52	0.9697	0.9812
norm | Matern 5/2	9	163	158	7586	6	0.51	0.9720	0.9839
innov | Matern 5/2	10	162	143	7601	7	0.53	0.9700	0.9813

Table A.7: Evaluation of covariance functions - VO attitude. Row captions specify the following parameter settings: Features (Feat.) {norm of innovations, innovation vector}, Covariance function {Squared Exponential (SE), Rational Quadratic (RQ), Matern $1/2, 3/2, 5/2$}

Feat. | Covariance function	#	TP	FP	TN	FN	PR	G	AUC
norm | SE	1	25	78	1499	2	0.24	0.9382	0.9794
innov | SE	2	17	133	1444	10	0.11	0.7593	0.7783
norm | RQ	3	25	78	1499	2	0.24	0.9382	0.9792
innov | RQ	4	22	178	1399	5	0.11	0.8502	0.8976
norm | Matern 1/2	5	25	78	1499	2	0.24	0.9382	0.9758
innov | Matern 1/2	6	25	86	1491	2	0.23	0.9356	0.9737
norm | Matern 3/2	7	25	78	1499	2	0.24	0.9382	0.9782
innov | Matern 3/2	8	24	109	1468	3	0.18	0.9096	0.9708
norm | Matern 5/2	9	25	78	1499	2	0.24	0.9382	0.9789
innov | Matern 5/2	10	25	211	1366	2	0.11	0.8956	0.9493

Table A.8: Evaluation of covariance functions - Laser attitude. Row captions specify the following parameter settings: Features (Feat.) {norm of innovations, innovation vector}, Covariance function {Squared Exponential (SE), Rational Quadratic (RQ), Matern $1/2, 3/2, 5/2$}

Feat. | Covariance function	#	TP	FP	TN	FN	PR	G	AUC
norm | SE	1	8	128	1468	0	0.06	0.9591	0.9929
innov | SE	2	8	160	1436	0	0.05	0.9486	0.9793
norm | RQ	3	8	128	1468	0	0.06	0.9591	0.9928
innov | RQ	4	8	200	1396	0	0.04	0.9352	0.9808
norm | Matern 1/2	5	8	129	1467	0	0.06	0.9587	0.9914
innov | Matern 1/2	6	8	314	1282	0	0.02	0.8962	0.9753
norm | Matern 3/2	7	8	129	1467	0	0.06	0.9587	0.9933
innov | Matern 3/2	8	8	239	1357	0	0.03	0.9221	0.9756
norm | Matern 5/2	9	8	128	1468	0	0.06	0.9591	0.9936
innov | Matern 5/2	10	8	211	1385	0	0.04	0.9316	0.9771

Table A.9: Evaluation of covariance functions - Laser position. Row captions specify the following parameter settings: Features (Feat.) {norm of innovations, innovation vector}, Covariance function {Squared Exponential (SE), Rational Quadratic (RQ), Matern $1/2, 3/2, 5/2$}

Uncertainty level	#	TP	FP	TN	FN	PR	G	Valid Data
1	1	163	159	7585	6	0.51	0.9720	1.00
0.9	2	160	44	7700	9	0.78	0.9702	0.98
0.8	3	155	27	7717	14	0.85	0.9560	0.96
0.7	4	149	16	7728	20	0.90	0.9380	0.95
0.6	5	140	12	7732	29	0.92	0.9095	0.93
0.5	6	131	10	7734	38	0.93	0.8799	0.91

Table A.10: Evaluation of uncertainty levels (uncertain set to negative) - VO attitude. Row caption specify the following parameter settings: Uncertainty level {what percentage of maximum uncertainty creates the data that are considered as valid}. Column Valid data indicates, what portion of data was left after application of the uncertainty threshold.

Uncertainty level	#	TP	FP	TN	FN	PR	G	Valid Data
1	1	25	78	1499	2	0.24	0.9382	1.00
0.9	2	18	35	1542	9	0.34	0.8074	0.91
0.8	3	17	25	1552	10	0.40	0.7872	0.83
0.7	4	13	21	1556	14	0.38	0.6893	0.71
0.6	5	11	16	1561	16	0.41	0.6350	0.49
0.5	6	10	13	1564	17	0.43	0.6061	0.12

Table A.11: Evaluation of uncertainty levels (uncertain set to negative) - Laser attitude. Row caption specify the following parameter settings: Uncertainty level {what percentage of maximum uncertainty creates the data that are considered as valid}. Column Valid data indicates, what portion of data was left after application of the uncertainty threshold.

Uncertainty level	#	TP	FP	TN	FN	PR	G	Valid Data
1	1	8	128	1468	0	0.06	0.9591	1.00
0.9	2	8	55	1541	0	0.13	0.9826	0.93
0.8	3	8	43	1553	0	0.16	0.9864	0.89
0.7	4	8	35	1561	0	0.19	0.9890	0.82
0.6	5	8	29	1567	0	0.22	0.9909	0.73
0.5	6	8	25	1571	0	0.24	0.9921	0.51

Table A.12: Evaluation of uncertainty levels (uncertain set to negative) - Laser position. Row caption specify the following parameter settings: Uncertainty level {what percentage of maximum uncertainty creates the data that are considered as valid}. Column Valid data indicates, what portion of data was left after application of the uncertainty threshold.