Improving Multimodal Data Fusion for Mobile Robots by Trajectory Smoothing

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

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

Abstract

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

Keywords: Field Robots, Sensor Fusion, Search and Rescue Robots

1. Introduction

For successful deployment of mobile robots to complex dynamically changing environments, such as those typical for Urban Search & Rescue (USAR), reliable localization is crucial. In modern mobile robots, a popular solution lies in the combination of proprioceptive sensors, usually in form of an integrated Inertial Navigation System (INS), that captures the body dynamics at high rate, and an external source of aiding, using either vision [1] or range

*Corresponding author
Email addresses: kubelka.vladimir@fel.cvut.cz (Vladimír Kubelka), reinstein.michal@fel.cvut.cz (Michal Reinstein), svobodat@fel.cvut.cz (Tomáš Svoboda)

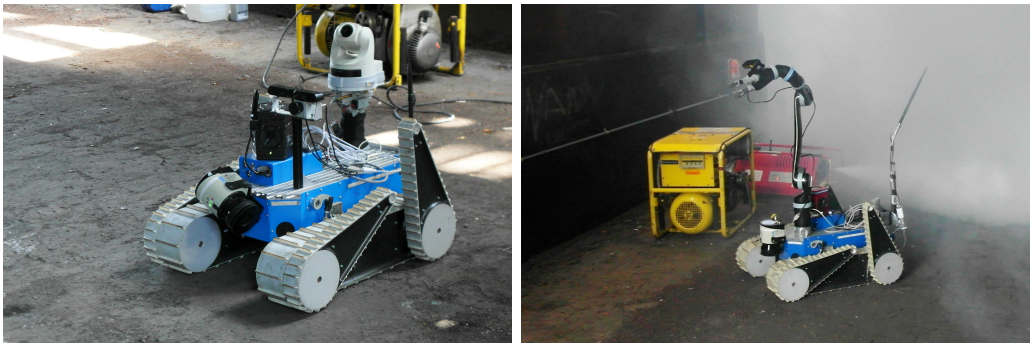

Figure 1: TRADR[2] robotic platform at the USAR training site of the fire brigade of Dortmund, Germany. Robot sensor suite is highly configurable, includes Point Grey Ladybug3 omni-directional camera and SICK LMS-151 laser range-finder, can be extended further by various cameras (left) or a robotic arm (right).

measurements [2]. Since most of the solutions are based on the well-proven Extended Kalman filter (EKF) [1, 2], the state estimation architecture we designed for our platform (see Figure 1) is based on the error state EKF framework as well.

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

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

the robustness of localization.

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

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

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

2. Related work

Regarding the multi-modal data fusion, we built on our previous results described in [3], especially the design of the EKF error models [9, 10, 11]—even though the later work concerned a legged robot.

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

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

Beside the visual odometry, another typical sensor for aiding is the laser range-finder. The laser range-finders are usually used for estimating vehicle motion by matching consecutive laser scans and thus creating a metric map of the environment [6, 7]. Examples of successful deployment can be found for indoor—without IMU but combined with vision [13]—as well as for outdoor—relying only on the IMU [2]. The most popular approach of scan matching is based on the Iterative Closest Point (ICP) algorithm, which was first proposed by [14] and [15]. Later, [16] proposed a 6D Simultaneous Localization and Mapping (SLAM) system relying primarily on the ICP. Work of [17] proposed a localization system combining a 2D laser SLAM with a 3D IMU/odometry-based navigation subsystem. Contrary to the later publications realized in the context of SLAM, we only consider the output of the ICP algorithm[4] as a local pose measurements—similarly as with the visual odometry, we treat the laser localization module as a *velocity sensor*.

Solutions exploiting the EKF for fusing the dead reckoning with exteroceptive sensors are very popular [1, 2, 19, 20, 21]; our fusion scheme is based on the EKF as well. Still, a number of problems arises in multimodal data fusion. The problem of utilizing several sensors for localization, which may provide contradictory measurements, is discussed in [22]. The authors use Bayes filters to estimate sensor measurement uncertainty and hence evaluate the sensor validity. We separately addressed this problem in [10], where we utilized machine-learning techniques to detect anomalous measurements.

Since we aim for grounding our approach with respect to the smoothers for Kalman filters (in order to smooth the trajectory estimates), we have chosen the well established RTS [8] for the EKF as our base reference for

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

benchmarking. Smoothers like RTS are well proven in the context of localization. In [23] a network of time-of-flight *Cricket* sensors provide measurements with a slight delay; the authors utilize an interacting multiple model fixed lag smoother to incorporate these delayed measurements. In [24] indoor localization problem is used to demonstrate properties of a smoother for the Unscented Kalman Filter. And finally in [25], the RTS smoother is actually utilized for the SLAM problem. Smoothing in Kalman filtering can be applied to wide range of problems, e.g. work of [26] applies the RTS to improve state estimation of a dynamic power system.

Several modifications of the RTS smoother have been proposed [27, 28, 29] since its original publication [8]. They mainly aim on better numerical stability and performance of the filter when deployed on computers with limited precision of number representation. We compare our algorithm to the RTS smoother using 64-bit double precision number representation and thus we take the liberty to use the original RTS formulation from [8].

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

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

3. Smoother for multimodal data fusion

3.1. Platform and sensors

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

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

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

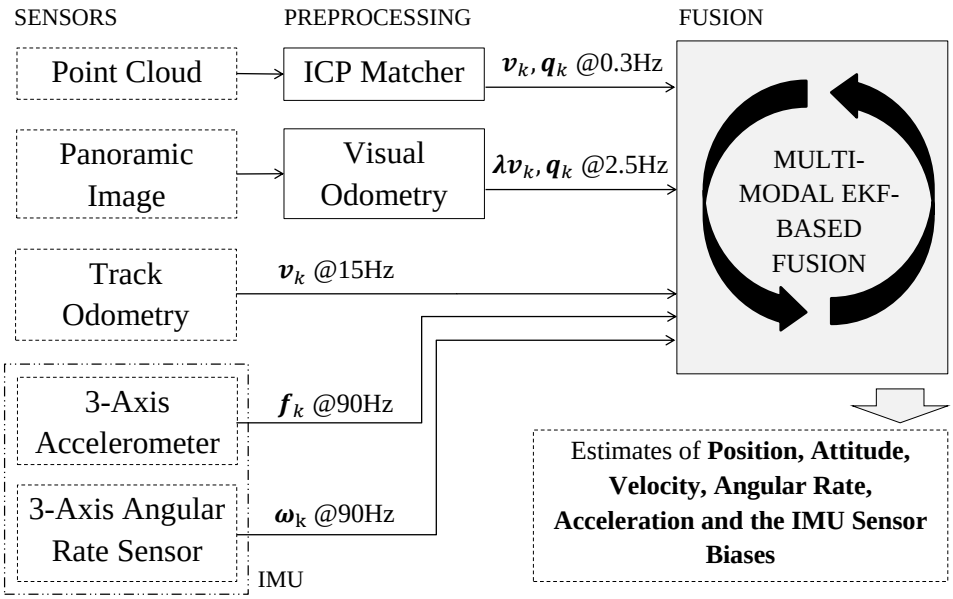

Figure 2: Schematics of the multi-modal fusion algorithm.

3.2. Multimodal data fusion framework

Output of the four modalities is preprocessed or directly enters the EKF-based[5] fusion algorithm that estimates—among others—the 3D position and orientation of the robot. Figure 2 presents the outline of the fusion system where $\boldsymbol{v}_k$ and $\boldsymbol{\omega}_k$ stand for linear and angular velocities, $\boldsymbol{f}_k$ is specific force measured by the accelerometers. Attitude provided by the laser range-finder and visual odometries is represented by quaternion $\boldsymbol{q}_k$; note that visual odometry indicates velocity multiplied by an unknown scale λ, which has to be estimated by the EKF fusion algorithm. As explained in [3], in order to design a modular solution, we treat all the modalities as velocity sensors.

The exteroceptive sensors generate measurements that require preprocessing in order to obtain information about translation and rotation of the robot. The laser range-finder creates 3D scans of the environment. We use *libpointmatcher* implementation [18] of the ICP algorithm to determine a

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

rigid transformation that would match two successive point clouds together. Resulting transformation reflects movement of the robot between the two 3D scans. According to our experience, this estimate is reliable in most environments the robot operates in. Its main limitation—the low sampling rate of one scan per 3 seconds—is given by the speed of rotation of the laser range-finder.

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

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

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

3.3. Scale-dependent velocity information from visual odometry

Relative motion of the robot is also measured by the visual odometry (VO). It benefits from the omni-directional camera observing the whole scene (up to occlusions caused by the robot body) and thus it is not loosing observed image features while turning. The omni-directional camera consists of several perspective cameras built in a way that purposively puts individual camera centers very close to each other and it is modeled as a central

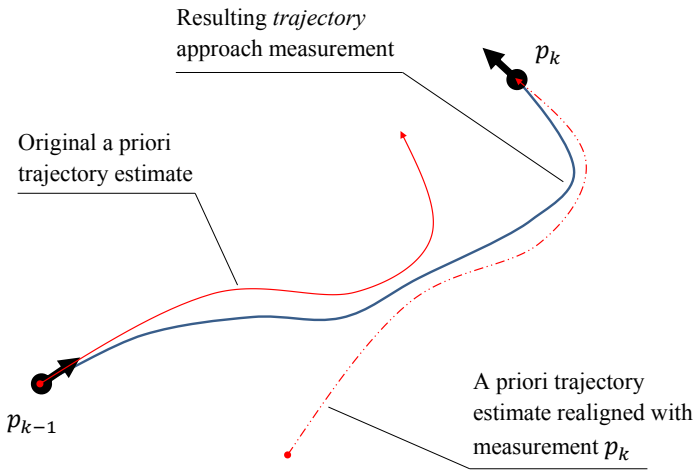

Figure 3: The trajectory approach [3]. Trajectory estimated from all modalities except the low-sampling-rate sensor (solid red line) is combined with position information (p_k,p_{k-1}) from the low-sampling-rate sensor (laser odometry). The new trajectory (thick blue) is a weighted average of the original trajectory and its aligned duplicate (red dot-dashed line).

omni-directional camera. Therefore, scale of the scene cannot be computed directly and it is up to the fusion algorithm to estimate it. The estimated scale then affects the velocity indicated by the VO. Since the VO scale has not been the part of the original fusion algorithm we proposed in [3] (and thus only VO attitude corrections could be utilized), we augment it to the state vector as follows:

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

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

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

Measurement model for velocity indicated by VO is then

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

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

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

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

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

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

Before the value of scale of the VO converges, invalid corrections of the velocity v_R are propagated into the system state. Similarly, erroneous track odometry measurements (e.g. in case of slippage) will eventually propagate into the estimate of λ. Therefore, to be able to utilize the visual odometry in a reasonable manner, we propose a modification that separates the λ estimation phase from the VO aiding phase by omitting either the $\hat{\lambda}\Delta\boldsymbol{v}_R$ term or the $\hat{\boldsymbol{v}}_R\Delta\lambda$ term from (4). In the first case, only λ is estimated without any v_R corrections being introduced. Based on the current EKF filter setting, we empirically determined it takes approximately 1.5 meters traveled by the robot for the λ to converge to a stable value (travel speed of the robot is in average 0.4m/s). After this nominal distance, we switch the terms, so from that point, λ is constant and every measurement residual is propagated into v_R correction. Of course, this approach expects the visual odometry to maintain constant scale, which may not always be the case[6].

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

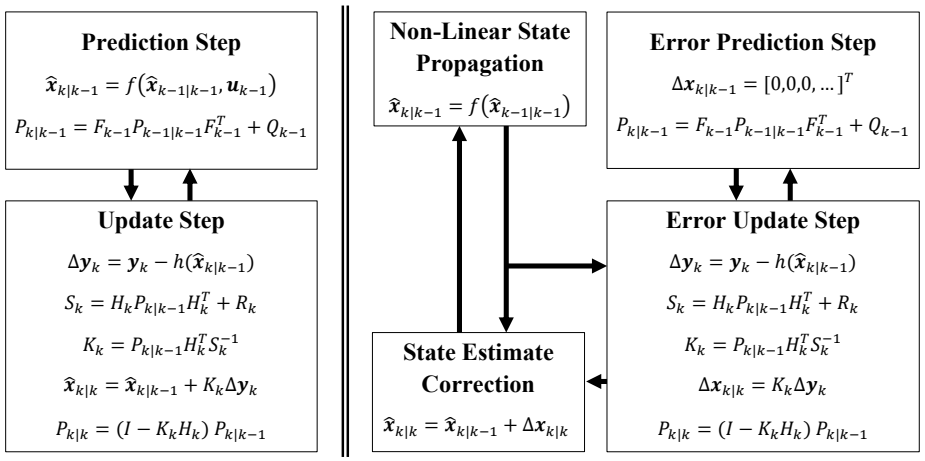

Figure 4: Comparison of the standard (left) and error-state (right) Extended Kalman filter computational scheme. Note that the system matrix F_k differs between the two filter implementations, it describes either dynamics of the system state or of the error of the system state [3]. Both F_k and H_k (measurement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

3.4. RTS smoother for the EKF

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

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

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

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

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

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

where the gain A_k is obtained from covariance matrices P and the system matrix F[7] as follows

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

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

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

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

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

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

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

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

4. Experimental evaluation

4.1. Dataset description

For the purpose of experimental evaluation, we used our dataset that we already released to the community[8]. It consists of experiments performed both outdoors and indoors, under reasonably challenging conditions—track slippages that occur naturally with the robot movement, traversing stairs,

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

[8] [3, 41], the dataset is recorded in a ROS *bagfile* format. Precise ground-truth position reference as captured by Vicon system is included.

slopes, gravel etc. We also performed series of experiments, where we deliberately diminished sensors performance (covered part of the omni-directional camera, scene over-exposure, artificially limited laser range-finder range) or simulated rough conditions (slippery slopes, deformable and unstable surfaces, hitting obstacles; see Fig. 5). The complete dataset contains approximately 3.2 km experiments with ground truth. To be able to evaluate performance, indoor experiments (total length 1.6km) were tracked by the Vicon system. With the ground truth, we can evaluate localization error in any given time yet such a metric is difficult to interpret, mainly because the error grows with distance traveled. To overcome this problem, we normalize the accumulated error by the distance and call it *average position error*:

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

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

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

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

4.2. Evaluation

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

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

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

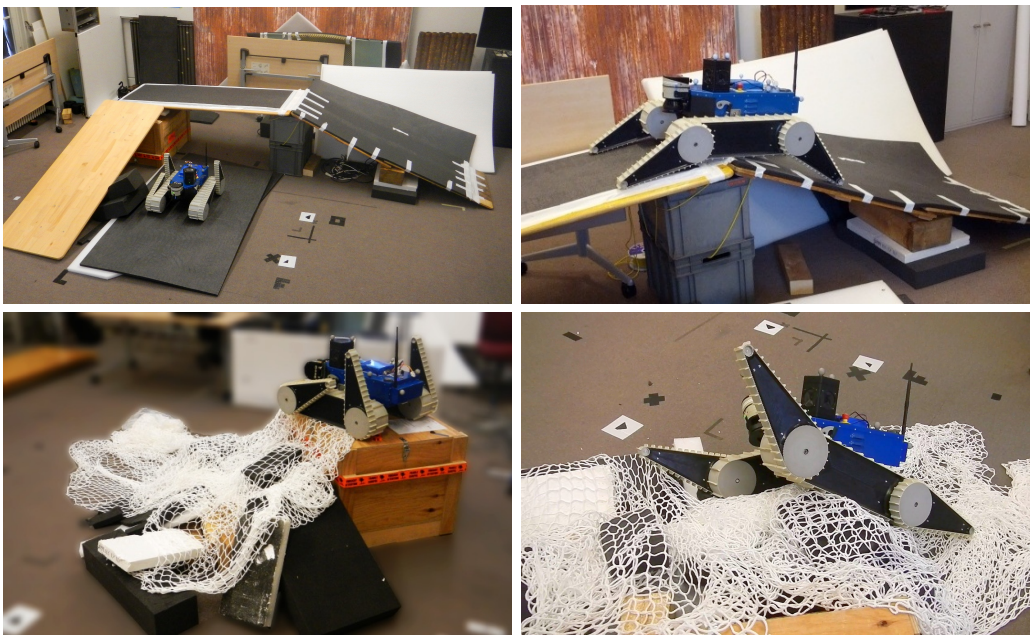

Figure 5: Examples of obstacles traversed during indoor localization experiments (Vicon tracking system as a ground truth). Top images: Ramp with a slippery surface (lacquered wood) and a non-slippery one (soft rubber). Bottom images: Pile of soft plastic foam, polystyrene and a nylon net (the wooden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

First exaple of experiments in the Fig. 6,7 compares the discussed approaches on a basic square trajectory. Both the incremental position approach and its RTS smoothed version show drift in the Z-axis, which is suppressed by the non-holonomic constraint in the case of the trajectory approach. The same effect of the trajectory approach is shown in a long corridor experiment in the Fig. 8. The reference system was a Leica Total Station[9].

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

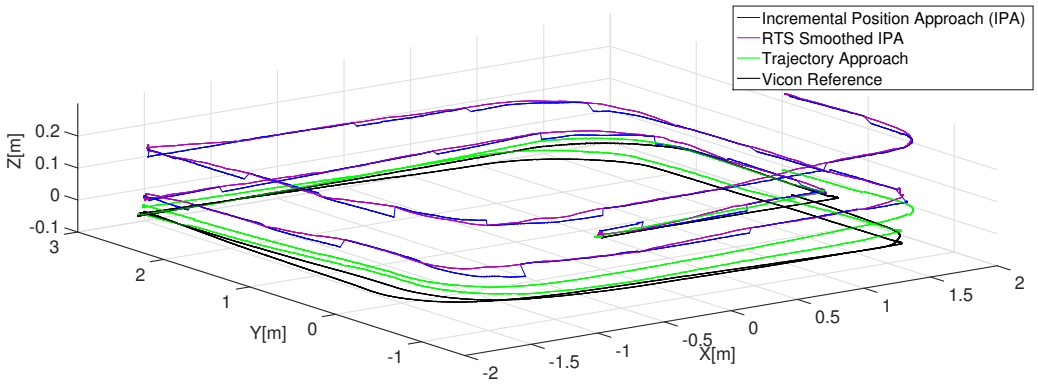

Figure 6: Comparison of the discussed approaches in an experiment with nominal conditions—robot was navigated in a square-shaped pattern on a leveled floor.

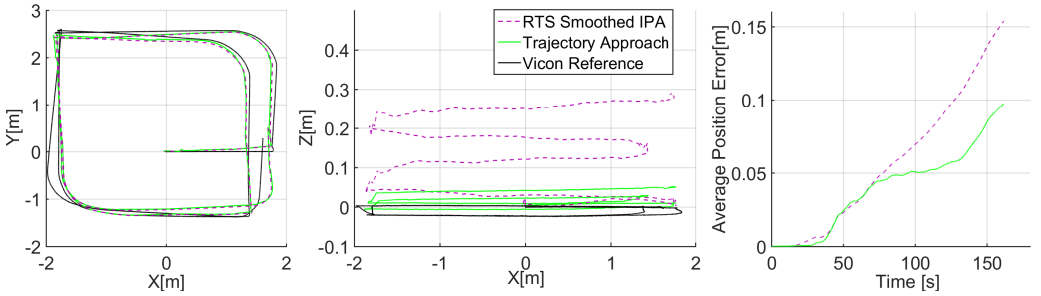

Figure 7: Effect of drift in the Z-axis for the RTS smoothed incremental position approach (IPA) compared to the trajectory approach (same experiment as in the Fig. 6).

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

An extreme fail-case of violating assumptions laid by the trajectory approach is demonstrated in the next example in the Fig. 10. The robot traversed two seesaws made of wooden pallets and large wooden blocks serving as a pivot. At some point of the traversal, the pallet together with the robot pivoted over while performing rotational and translational motion. Both

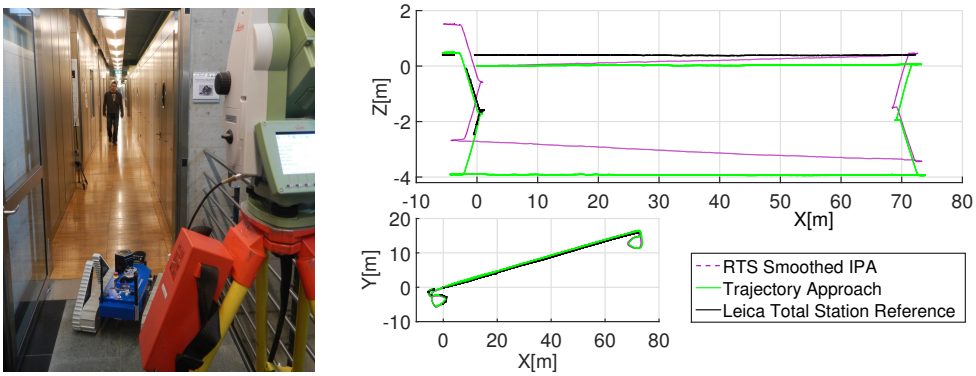

Figure 8: Effect of the discussed approaches in a long narrow corridor experiment spanning two floors. Non-holonomic constraints in measured velocity (allowed by the trajectory approach) reduced the Z-axis drift. Note that the reference system was available only on the second floor and the offset between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

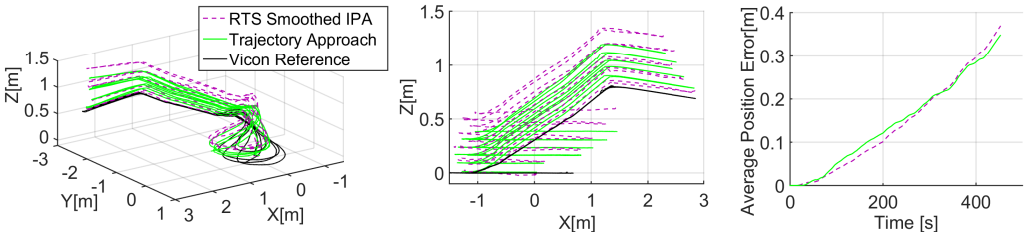

Figure 9: A ramp-climbing fail-case experiment example. The robot repeatedly climbed an inclined ramp (see Fig. 5, top right). With each pass, error in the Z coordinate estimate increases because of inaccurate readings of the tracked odometry when rolling onto the inclined ramp and pivoting over its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

of motion components were sensed by the inertial measurement unit. Rotational part was estimated without any problems, however, the translation was based on data from visual odometry, IMU acceleration and track odometry which provided contradictory measurements. This resulted into incorrect localization for both approaches; non-holonomic constraints were violated.

The last presented fail-case related to the low-rate ICP odometry is depicted in the Fig. 11. In this experiment, we intentionally paused laser range finder for up to 45 seconds simulating ICP odometry outages. During these outages, the robot passed over slippery and non-slippery ramp inducing track odometry errors discussed in previous experiment examples. It is necessary to comment on the middle plot in the Fig. 11: Compared to our fused lo-

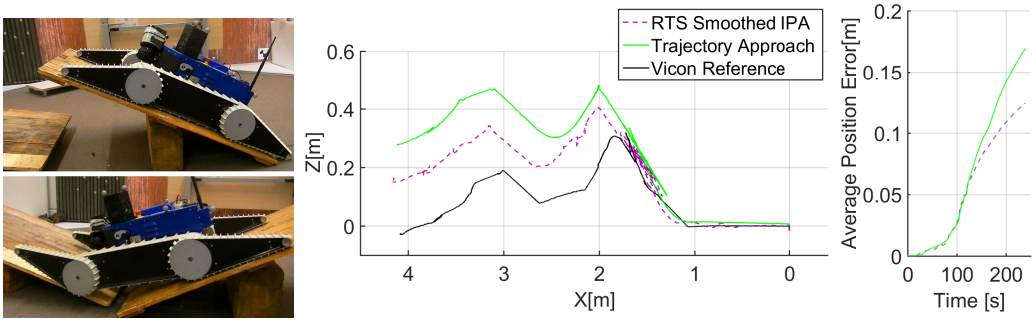

Figure 10: An extreme case with slippage and swinging pallets. The pivoting motion of the pallet lying on the wooden block violates the non-holonomic constraints used in the trajectory approach. That results in worse position estimate than the smoothed incr. position approach.

calization output, the bare ICP odometry yields minimal error. That is expected in a laboratory where it has perfect conditions for localization—six large planes of walls, ceiling and floor. This advantage does not hold in general, ICP odometry tends to perform poorly in uniform tube-shaped environments (e.g. long corridors). Nevertheless, comparison of the smoothed incremental position approach and the trajectory approach favors the first one. While slippage does not usually significantly deteriorate the trajectory approach performance, the long outages lead to poor trajectory reconstruction (Fig. 3), especially in the Z-axis leading to stronger drift[10].

4.3. Comparative analysis

In the Table 1, we evaluate performance of the approaches on the indoor set of experiments—the Vicon reference system provides higher sampling rate and thus allows us to evaluate overall accuracy as well as details of the estimated robot trajectory. The first column contains the set of experiments with nominal conditions; in the second column, the fail-case experiments are also included. The results show that the overall accuracy is comparable—the median values of average position error of each approach is well within lower and higher quartiles of the other approaches. We choose to express the results by lower and higher quartile and median since our metric value

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

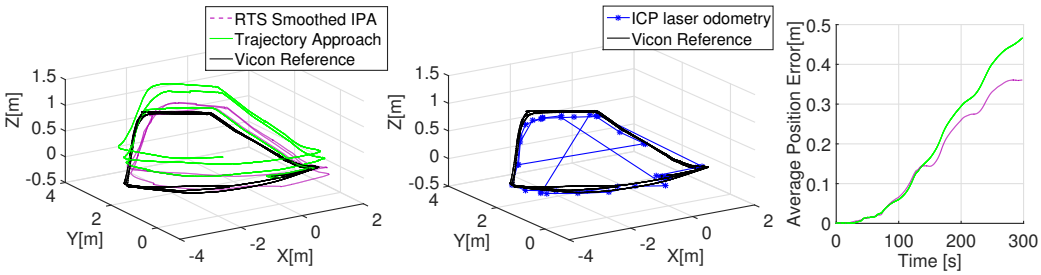

Figure 11: An ICP laser odometry outage fail-case. The raw ICP measurements are shown in the middle plot (blue stars are the points of the ICP measurements); the nominal rate is 0.3 Hz however in this experiment, we switched laser for longer periods to simulate laser failures. The robot path also included the slippery ramp that caused invalid velocity measurements from track odometry.

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

Approach	Normal conditions (777m,7218s)	+ Hard conditions (2174m,15931s)
incremental position	0.08|**0.12**|0.22	0.10|**0.18**|0.39
incr. pos. with smoothing	0.08|**0.12**|0.21	0.10|**0.19**|0.38
trajectory approach	0.11|**0.14**|0.24	0.12|**0.19**|0.36
	lower|**median**|higher	quartile [m]

is not normally distributed. Including the fail-cases increases the median of the average position error, as expected.

4.4. Discussion

With respect to the presented results, obvious question arises: which approach should be chosen? On flat or reasonably inclined surfaces without rough transitions, trajectory approach thanks to non-holonomic constraint on ICP odometry velocity measurements corrects localization drift (predominantly in the Z-axis). On rough and deformable terrain and when pivoting over obstacle edges, the non-holonomic constraint does not hold anymore and therefore, the standard incremental position approach performs similarly to our trajectory approach. In the Vicon-referenced part of our dataset, the dominant part are experiments where the robot crosses ramps and other

obstacles. Therefore, the comparison of the approaches in the Table 1 indicates similar performance. The choice of the approach therefore depends on the expected conditions of the robot deployment. If a significant part of trajectories driven by the robot contains large flat areas (structured environment in general), the localization can benefit from non-holonomic constraints implemented in the proposed trajectory approach.

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

5. Conclusions

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

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

Acknowledgment

The research presented here was supported by the European Union FP7 Programme under the TRADR project (No. 609763; http://www.tradr-project.eu), by the Czech Science Foundation (Project Registration No. 14-13876S) and by The Grant Agency of the CTU Prague under Project SGS15/081/OHK3/1T/13.

We would like to thank F. Colas, F. Pomerleau and L. Oswald for their valuable comments and advices and for their assistance with creation of the experimental dataset.

References

[1] G. Chowdhary, E. N. Johnson, D. Magree, A. Wu, A. Shein, GPS-denied Indoor and Outdoor Monocular Vision Aided Navigation and Control of Unmanned Aircraft, Journal of Field Robotics 30 (3) (2013) 415–438. doi:10.1002/rob.21454.
URL http://dx.doi.org/10.1002/rob.21454

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

[3] V. Kubelka, L. Oswald, F. Pomerleau, F. Colas, T. Svoboda, M. Reinstein, Robust Data Fusion of Multimodal Sensory Information for Mobile Robots, Journal of Field Robotics 32 (4) (2015) 447–473. doi:10.1002/rob.21535.
URL http://dx.doi.org/10.1002/rob.21535

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

[5] D. Scaramuzza, F. Fraundorfer, Visual Odometry [Tutorial], IEEE Robot Autom Mag 18 (4) (2011) 80–92.

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

[7] T. Yoshida, K. Irie, E. Koyanagi, M. Tomono, A sensor platform for outdoor navigation using gyro-assisted odometry and roundly-swinging 3D laser scanner, in: Proc. IEEE/RSJ Int Intelligent Robots and Systems (IROS) Conf, 2010, pp. 1414–1420.

[8] H. E. Rauch, C. T. Striebel, F. Tung, Maximum likelihood estimates of linear dynamic systems, AIAA Journal 3 (8) (1965) 1445–1450.

[9] J. Simanek, M. Reinstein, V. Kubelka, Evaluation of the EKF-Based Estimation Architectures for Data Fusion in Mobile Robots, Mechatronics, IEEE/ASME Transactions on 20 (2) (2015) 985–990. doi:10.1109/TMECH.2014.2311416.

[10] J. Simanek, V. Kubelka, M. Reinstein, Improving multi-modal data fusion by anomaly detection, Autonomous Robots 39 (2) (2015) 139–154. doi:10.1007/s10514-015-9431-6.
URL http://dx.doi.org/10.1007/s10514-015-9431-6

[11] M. Reinstein, M. Hoffmann, Dead Reckoning in a Dynamic Quadruped Robot Based on Multimodal Proprioceptive Sensory Information, IEEE Transactions on Robotics 29 (2) (2013) 563–571. doi:10.1109/TRO.2012.2228309.

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

[13] L.-P. Ellekilde, S. Huang, J. V. Miro, G. Dissanayake, Dense 3D Map Construction for Indoor Search and Rescue, J. Field Robotics 24 (1-2) (2007) 71–89.

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

[15] Y. Chen, G. Medioni, Object modeling by registration of multiple range images, in: Robotics and Automation (ICRA), 1991. Proceedings of the IEEE International Conference on, 1991, pp. 2724–2729.

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

[17] S. Kohlbrecher, O. V. Stryk, J. Meyer, U. Klingauf, A flexible and scalable SLAM system with full 3D motion estimation, in: Safety, Security, and Rescue Robotics (SSRR), 2011 IEEE International Symposium on, 2011, pp. 155–160.

[18] F. Pomerleau, F. Colas, R. Siegwart, S. Magnenat, Comparing ICP Variants on Real-World Data Sets, Autonomous Robots 34 (3) (2013) 133–148.

[19] T. Oskiper, H.-P. Chiu, Z. Zhu, S. Samarasekera, R. Kumar, Multi-modal sensor fusion algorithm for ubiquitous infrastructure-free localization in vision-impaired environments, in: Intelligent Robots and Systems (IROS), 2010 IEEE/RSJ International Conference on, 2010, pp. 1513–1519. doi:10.1109/IROS.2010.5649562.

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

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

[22] S. R. Sukumar, H. Bozdogan, D. L. Page, A. F. Koschan, M. A. Abidi, Sensor Selection Using Information Complexity for Multi-sensor Mobile Robot Localization, in: Robotics and Automation, 2007 IEEE International Conference on, 2007, pp. 4158–4163.

[23] H. Song, V. Shin, Localization Using Multisensor Fusion of IMM Fixed Lag Smoother in a Cricket Sensor Network, in: Systems and Networks Communications (ICSNC), 2010 Fifth International Conference on, 2010, pp. 130–135. doi:10.1109/ICSNC.2010.26.

[24] D. Lee, M. Campbell, Iterative smoothing approach using Gaussian mixture models for nonlinear estimation, in: Intelligent Robots and Systems (IROS), 2012 IEEE/RSJ International Conference on, 2012, pp. 2498–2503. doi:10.1109/IROS.2012.6385752.

[25] S. Razali, K. Watanabe, S. Maeyama, K. Izumi, An unscented Rauch-Tung-Striebel smoother for SLAM problem, in: SICE Annual Conference (SICE), 2011 Proceedings of, 2011, pp. 1304–1308.

[26] J. Sreenath, S. Chakrabarti, A. Sharma, Implementation of Rauch-Tung-Striebel smoother for power system dynamic state estimation in

the presence of PMU measurements, in: Smart Grid Technologies-Asia (ISGT ASIA), 2015 IEEE Innovative, IEEE, 2015, pp. 1–6.

[27] S. R. McReynolds, Fixed interval smoothing-Revisited, Journal of Guidance, Control, and Dynamics 13 (5) (1990) 913–921.

[28] G. Bierman, A new computationally efficient fixed-interval, discrete-time smoother, Automatica 19 (5) (1983) 503–511.

[29] K. Watanabe, S. Tzafestas, New computationally efficient formula for backward-pass fixed-interval smoother and its UD factorisation algorithm, in: Control Theory and Applications, IEE Proceedings D, Vol. 136, IET, 1989, pp. 73–78.

[30] V. Indelman, S. Williams, M. Kaess, F. Dellaert, Factor graph based incremental smoothing in inertial navigation systems, in: Information Fusion (FUSION), 2012 15th International Conference on, IEEE, 2012, pp. 2154–2161.

[31] H.-P. Chiu, S. Williams, F. Dellaert, S. Samarasekera, R. Kumar, Robust vision-aided navigation using Sliding-Window Factor graphs, in: Robotics and Automation (ICRA), 2013 IEEE International Conference on, 2013, pp. 46–53.

[32] M. Kaess, H. Johannsson, R. Roberts, V. Ila, J. J. Leonard, F. Dellaert, iSAM2: Incremental smoothing and mapping using the Bayes tree, The International Journal of Robotics Research (2011) 0278364911430419.

[33] S. Lange, N. Sunderhauf, P. Protzel, Incremental smoothing vs. filtering for sensor fusion on an indoor UAV, in: Robotics and Automation (ICRA), 2013 IEEE International Conference on, IEEE, 2013, pp. 1773–1778.

[34] G. Sibley, L. Matthies, G. Sukhatme, Sliding window filter with application to planetary landing, Journal of Field Robotics 27 (5) (2010) 587–608.

[35] S. Leutenegger, S. Lynen, M. Bosse, R. Siegwart, P. Furgale, Keyframe-based visual–inertial odometry using nonlinear optimization, The International Journal of Robotics Research 34 (3) (2015) 314–334.

[36] D. G. Kottas, S. I. Roumeliotis, An iterative Kalman smoother for robust 3D localization on mobile and wearable devices, in: Robotics and Automation (ICRA), 2015 IEEE International Conference on, IEEE, 2015, pp. 6336–6343.

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

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

[39] P. G. Savage, Strapdown Inertial Navigation Integration Algorithm Design Part 2: Velocity and Position Algorithms, Journal of Guidance, Control, and Dynamics 21 (No. 2) (1998) 208 – 221.

[40] W. G. Breckenridge, Quaternions - Proposed Standard Conventions, Tech. rep., JPL (1999).

[41] V. Kubelka, F. Pomerleau, L. Oswald, M. Reinstein, NIFTi@Zurich dataset (2013).
URL `https://sites.google.com/site/kubelvla/public-datasets/nifti-zurich-2013`