

# Relative Topometric Localization in Globally Inconsistent Maps

Mladen Mazuran, Federico Boniardi, Wolfram Burgard, and Gian Diego Tipaldi

**Abstract** Mobile robot localization is a mature field that over the years has demonstrated its effectiveness and robustness. The majority of the approaches, however, rely on a globally consistent map, and localize on it in an absolute coordinate frame. This global consistency cannot be guaranteed when the map is estimated by the robot itself, and an error in the map will likely result in the failure of the localization subsystem. In this paper we introduce a novel paradigm for localization, namely *relative topometric localization*, by which we forgo the need for a globally consistent map. We adopt a graph-based representation of the environment, and estimate both the topological location on the graph and the relative metrical position with respect to it. We extensively evaluated our approach and tested it against Monte Carlo localization on both simulated and real data. The results show significant improvements in scenarios where there is no globally consistent map.

## 1 Introduction

Mobile robot localization is a well studied field in robotics and several robust approaches to localization have been proposed in the past [11, 14, 19, 12]. The majority of those approaches, however, assume that a globally consistent map of the environment is built beforehand by a mapping process. This map is then used to estimate the pose of the robot in a single absolute reference frame, often without taking into account the map uncertainties arising from the robot pose estimates during mapping.

Such maps are normally built by a separate mapping process, solving what is called the simultaneous localization and mapping (SLAM) problem. Many modern techniques for this problem are based on a least squares minimization approach over a graph of measurements [13], whose error is assumed to be normally distributed. In the presence of data association outliers, the Gaussian error assumption leads

---

Mladen Mazuran, Federico Boniardi, Wolfram Burgard, and Gian Diego Tipaldi  
University of Freiburg, 79110, Freiburg, Germany, e-mail: {surname}@cs.uni-freiburg.de

to maps that are not globally consistent. The need of a globally consistent map for localization led many researchers to either apply robust statistics in the minimization step [1, 30, 18, 24], or to measure the global consistency of the resulting maps [20].

In this paper we propose a novel paradigm to robot localization that relaxes the assumption of a globally consistent map and a single absolute reference frame. We believe that, for the majority of navigation tasks, a globally consistent map is not necessary and one only needs global topological information and local metrical consistency. Our paradigm, *relative topometric localization* (RTL), is based on a graph representation of the environment, where each node represents a pose in the map and each edge their relative transformation. The graph implicitly defines a manifold structure with a chart associated to each node that parametrizes its local neighborhood. Such maps are not unusual in robotics, since they are the internal representation of modern mapping approaches based on pose-graph estimation or bundle adjustment. We also do not assume this map to be the result of any optimization process, nor to be globally embeddable in a single Euclidean space.

We formulate the localization problem as jointly estimating the current reference frame and the relative pose of the robot within its chart. Our paradigm has a set of advantages with respect to approaches based on a single absolute reference frame: 1) it is inherently robust to maps that are not globally consistent; 2) it includes uncertainties in the map estimate during localization; and 3) it operates on unoptimized maps, removing the need for a SLAM back-end.

We thoroughly evaluated our approach on a large set of experiments in fully simulated environments, environments sourced from real data, as well as on a real robot tracked with a motion capture system. To verify our claims, we compared our method with a Monte Carlo localization approach [26] on maps that have different levels of global inconsistency. In an additional experiment, we also evaluated our paradigm when the map is globally consistent but the environment changed its appearance (e.g., furniture was added or removed). The experimental results demonstrate that a relative topometric approach is indeed resilient to global inconsistencies in the map. Further, the method provides localization accuracy in the order of millimeters even under substantial changes in the environment or inconsistencies.

## 2 Related Work

Autonomous localization has been mainly addressed within probabilistic state estimation frameworks, with solutions based on extended Kalman filters (EKF) [19], histogram filters [12] or particle filters, often referred to as Monte-Carlo localization (MCL) [11]. Such approaches assume knowledge of a globally consistent map, without considering any uncertainty on it apart from the error induced by grid-based approximations. In this work, we relax the assumption of global consistency and we explicitly consider uncertainties in the map stemming from the SLAM process.

In the context of SLAM and bundle adjustment, several authors explored the concept of relative estimates. Howard *et al.* [2] introduced the idea of a manifold

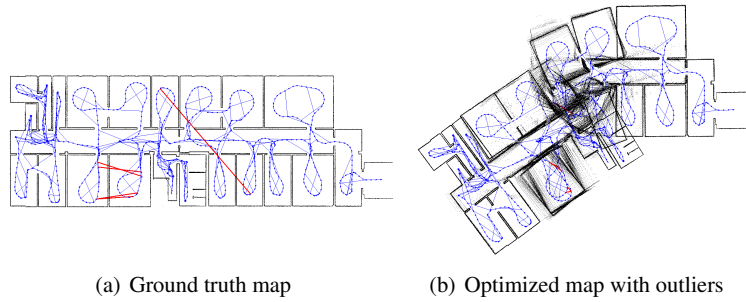
mapping for multiple robots. In their approach, the map is represented as a two-dimensional manifold embedded in a higher-dimensional space. They introduce key ideas of the manifold structure and present an application to multi-robot mapping. Sibley *et al.* [27] proposed the relative bundle adjustment paradigm. They claim that bundle adjustment is costly due to the choice of a single privileged reference frame and propose to optimize in a metric space defined by a manifold instead of a single Euclidean space. Their paradigm has been first extended by Blanco *et al.* [6], to consider a set of possible sparsification strategies, and by Strasdat *et al.* [29], to consider a second optimization window and to enforce a metric consistency within this optimization window. We differ from those works for the fact that we address a localization problem and also estimate a distribution over the reference frame. This is needed in order to include possible multiple localization hypotheses.

Churchill and Newman [9] introduced the concept of *navigation experiences*, i.e., robot paths with relative metrical information. They localize the robot in the experiences by first using appearance-based data association methods to estimate the initial node in the experience graph and then track the robot position using visual odometry techniques. In contrast to our framework, their approach does not consider uncertainties in the map, nor does it track the index of the reference frame over time.

Recently, the need of a globally consistent map for navigation has been questioned by some researchers. Konolige *et al.* [15] proposed a navigation system based on a hybrid metric-topological map. They employ a laser scanner and localize the robot with respect to one reference node in the graph. Dayoub *et al.* [10] extended the approach to cameras and considered a set of image features associated at each node in the graph. Our approach is closely related to these last two. The main difference is that we do not assume the graph to be the result of an optimization algorithm. Additionally, we are able to consider uncertainties in the map estimates.

The concept of topometric localization has also been exploited by Badino *et al.* [3]. The authors build a topological map as a vehicle travels along its route. Each node is linked with a pose in the environment and the localization algorithm estimates which node the robot is currently in. The robot pose is then the pose stored in that node. Xu *et al.* [33] extended this to consider multiple roads and branchings along paths. Both approaches differ from ours, as the real estimation part is purely topological and the authors still rely on a global reference frame for the robot pose.

Ideas closely related to topometric localization are also present in teach-and-repeat frameworks. Sprunk *et al.* [28] proposed a data-driven approach using laser scanners and demonstrated millimeter-level accuracy without building any globally consistent map of the environment. McManus *et al.* [21] proposed an approach based on 3D laser reflectivity that is able to handle long-range navigation in challenging environments. The map is a chain of poses, where each pose has a submap associated to it. Localization is performed on the latter, relative to the corresponding pose, and a set of heuristics to switch submaps is presented. Krüsi *et al.* [16] build a similar metrical/topological map, but rely on ICP to perform localization on the submaps with a local motion planner to avoid obstacles. Our approach differs from teach and repeat paradigms since we do not localize only on a single route and we further propose a sound estimation framework for tracking the reference frame.



**Fig. 1** Effect of outliers in the association on the global consistency of the resulting map. Introducing just 5 outliers in the data association (marked in red) renders the central part of the map completely unusable for MCL.

### 3 Relative Topometric Localization

In this section we describe the *relative topometric localization* paradigm and relate it to the *metric* one. In metric localization, one seeks to estimate the pose of the robot at each time step, given the map of the environment and the history of sensor measurements and controls. The main assumption is that the map is a globally consistent metric representation of the environment, consisting of the absolute position of relevant features in the environment (e.g., cells in an occupancy grid or a set of key-points for visual localization). The location of those features, as well as the pose of the robot, is expressed with respect to a single reference frame, which is the global coordinate frame of the map. Formally, this is equivalent to recursively estimating the posterior  $p(\mathbf{x}_t \mid \mathbf{z}_{1:t}, \mathbf{u}_{1:t}, m)$ , with robot pose  $\mathbf{x}_t$ , measurements  $\mathbf{z}_{1:t}$ , odometry readings  $\mathbf{u}_{1:t}$ <sup>1</sup> and map  $m$ .

This is often estimated by a finite set of particles, resulting in the well established field of Monte Carlo localization (MCL) [11]. MCL has been proven to be a robust approach for localization and is often deployed on robots when the map  $m$  is known a-priori. Nevertheless, estimating the pose of the robot in an absolute frame of reference implicitly assumes the map  $m$  to be globally consistent, an assumption that can easily be violated in the presence of outlier associations during the SLAM process. Such a scenario can especially occur if the map of the environment is automatically computed by the robot, without human post-processing, and the underlying SLAM system introduces wrong associations due to, say, perceptual aliasing [1]. Fig. 1 showcases how the effect of a small amount of associations can impact the resulting map and consequently cripple the estimate of MCL in a large portion of it.

<sup>1</sup> Note that this is equivalent to the more traditional formulation with control inputs.

### 3.1 Relative Topometric Paradigm

In the relative topometric paradigm, we relax the assumption of a globally consistent metric map. As map representation, we follow an approach much akin to the one proposed by Howard *et al.* [2]. We consider the map of the environment to be a collection of patches that are locally homeomorphic to the Euclidean space, thus inducing a manifold structure, without the need of defining a global embedding.

More precisely, the map is a graph of poses  $\mathbf{m}_i$ ,  $i = 1, \dots, N$  with a set of relative transformations  $\mathbf{w}_{i,j} \in \text{SE}(n)$  between them, and sensory data attached to them. Each node  $i$  in the graph defines a local chart of the manifold, of which  $\mathbf{m}_i$  is the origin. Here, we can assume that the projection of the sensory data of the nodes in a neighborhood of  $i$  is roughly locally consistent. This, in turn, defines an open set in which the chart is valid.

While the chart is only locally valid, we can still express the poses of the remainder of the map in this reference system by chaining the transformations  $\mathbf{w}_{i,j}$  along the minimum distance spanning tree of the whole graph, with farther nodes having increasingly erroneous relative position estimates.

Let  $\mathbf{z}_t = [\mathbf{z}_t^{(1)} \mathbf{z}_t^{(2)} \dots \mathbf{z}_t^{(N)}]$  be a vector of measurements at time  $t$ , where each  $\mathbf{z}_t^{(i)}$  arises from matching the current observations of the robot with the observations stored in the node  $i$  of the graph. We wish to estimate at each time step  $t$  the vector of relative transformations between the poses  $\mathbf{m}_i$  and the robot, namely  $\Delta \mathbf{x}_t = [\Delta \mathbf{x}_t^{(1)} \Delta \mathbf{x}_t^{(2)} \dots \Delta \mathbf{x}_t^{(N)}] \in \text{SE}(n)^N$ . In the ideal situation, where the manifold can be globally embedded,  $\Delta \mathbf{x}_t^{(i)} = \ominus \mathbf{m}_i \oplus \mathbf{x}_t$ . Here,  $\mathbf{m}_i$  and  $\mathbf{x}_t$  respectively represent the pose of the node and of the robot in an absolute coordinate frame.

Formally, this is equivalent to jointly estimating the posterior over the reference frame  $r_t$  (i.e., the index of the current chart for the manifold) and the relative poses  $\Delta \mathbf{x}_t$  of the robot in that particular chart:

$$p(r_t, \Delta \mathbf{x}_t \mid \mathbf{z}_{1:t}, \mathbf{u}_{1:t}, \mathbf{w}) = p(\Delta \mathbf{x}_t \mid r_t, \mathbf{z}_{1:t}, \mathbf{u}_{1:t}, \mathbf{w}) p(r_t \mid \mathbf{z}_{1:t}, \mathbf{u}_{1:t}, \mathbf{w}). \quad (1)$$

The distribution is composed of a discrete probability mass, associated with a continuous probability density function, which we assume to be a Gaussian. The former represents the probability of the robot being within the open set in which the chart is valid, the latter represents the distribution of the robot pose in that specific chart. Following Bar-Shalom *et al.* [4], the estimation of (1) can be formulated within the *Dynamic Multiple Model Estimator* framework. In this framework, the goal is to estimate the state of a system that obeys one of a finite number of system models. Each model defines its own measurement and motion model and, during the estimation process, the system may change its model of operation. In our case, we can consider the choice of the reference frame as one model of our system, resulting in a finite number of models equivalent to the number of nodes in our graph. The change of model happens when the robot is not any longer in the valid set of the current chart and a new chart must be used.

Unfortunately, the full estimation problem requires tracking the whole sequence of frames, which results in Gaussian mixture distribution with an exponentially increasing number of terms. To reduce this complexity, approximated algorithms have been proposed, such as the Interacting Multiple Model [4], Markov Chain Monte Carlo [23], or Multiple Hypothesis Tracking [5].

In this paper, we focus on position tracking and we approximate the estimation of (1) by considering only the maximum likelihood frame at each time step. We believe that this is a valid approximation for position tracking problems, which is also confirmed by the experimental results presented below. We leave the full estimation problem as future work.

## 4 Position Tracking in the Relative Topometric Paradigm

This section describes the proposed approximation of (1) and its instantiation for position tracking. At each time step  $t$ , we approximate the distribution over the reference frames with a deterministic distribution, centered at its maximum  $\hat{r}_t$ , i.e.,

$$p(r_t | \mathbf{z}_{1:t}, \mathbf{u}_{1:t}, \mathbf{w}) \approx \delta[r_t - \hat{r}_t], \quad (2)$$

where  $\delta[r]$  denotes the discrete impulse function. To compute the next estimate of  $r_t$  we thus need to find the index  $i$  over the possible reference frames that maximizes

$$p(r_t = i | \mathbf{w}, \mathbf{z}_{1:t}, \mathbf{u}_{1:t}) = \sum_{r_{t-1}} p(r_t = i | r_{t-1}, \mathbf{z}_{1:t}, \mathbf{u}_{1:t}, \mathbf{w}) p(r_{t-1} | \mathbf{z}_{1:t}, \mathbf{u}_{1:t}, \mathbf{w}) \quad (3)$$

$$\propto p(\mathbf{z}_t | r_t = i, \hat{r}_{t-1}, \mathbf{u}_{1:t}, \mathbf{w}) p(r_t = i | \hat{r}_{t-1}, \mathbf{z}_{1:t-1}, \mathbf{u}_{1:t}, \mathbf{w}) \quad (4)$$

$$\approx p(\mathbf{z}_t | r_t = i, \hat{r}_{t-1}, \mathbf{u}_t, \mathbf{w}, \Delta \hat{\mathbf{x}}_{t-1}) p(r_t = i | \hat{r}_{t-1}). \quad (5)$$

Here,  $\Delta \hat{\mathbf{x}}_{t-1}$ , denotes the random variable associated to the posterior distribution of  $\Delta \mathbf{x}_{t-1}$  estimated at the previous time step. We assume  $\Delta \hat{\mathbf{x}}_{t-1}$  to be approximable by a Gaussian random variable with mean  $\Delta \bar{\mathbf{x}}_{t-1}$  and covariance  $\boldsymbol{\Sigma}_{t-1}$ . Note that from (3) to (4), we substituted approximation (2), applied Bayes' rule on  $\mathbf{z}_t$ , and used the independence assumption of the measurements. From (4) to (5) we followed the same approach of the *generalized pseudo-Bayesian estimator of first order* (GPB1) [4], namely, we assumed a transition model  $p(r_t | r_{t-1})$  on the reference nodes, and used the previous displacement estimate for computing the measurement likelihood.

In this work we restrict ourselves to a uniform  $p(r_t | r_{t-1})$  over all nodes in a local neighborhood of  $r_{t-1}$ , although this can be easily generalized to more elaborate models. As for the measurement likelihood, we computed it as follows. We first predict the relative pose of the robot in the reference frame  $r_t = i$  by propagating the previous relative pose at time  $t - 1$  from the reference  $\hat{r}_{t-1}$  through the relative measurements  $\mathbf{w}$  defined in the map. Under normality assumptions, the likelihood is then equal to  $\mathcal{N}(\mathbf{z}_t - \hat{\mathbf{z}}_t; \mathbf{0}, \boldsymbol{\Sigma})$ , where  $\hat{\mathbf{z}}_t$  is the predicted observation relative to the reference frame  $r_t = i$  and  $\boldsymbol{\Sigma}$  is the innovation covariance matrix.

In the case of range sensors, which is the implementation target of this paper, the map includes a laser scan per node, the predicted observation is the relative pose of the robot in the reference frame of  $r_t$ , while the current observation can be the result of an ICP procedure [8].

Let  $\mathbf{m}$  denote the stacked  $\mathbf{m}_i$  poses, in the local neighborhood of  $r_t$ , say  $V_t$ , for which the chart associated to  $r_t$  is valid, as well as the pose  $\mathbf{m}_s$  with  $s = \hat{r}_{t-1}$ . Then, to compute the distribution over the robot pose, we express (1) as the marginal over  $\mathbf{m}$  and  $\Delta \mathbf{x}_{t-1}$ . Given the approximation defined in (2), we compute the distribution

$$\begin{aligned} & p(\Delta \mathbf{x}_t \mid \hat{r}_t, \mathbf{z}_{1:t}, \mathbf{u}_{1:t}, \mathbf{w}) \\ &= \iint p(\Delta \mathbf{x}_t, \Delta \mathbf{x}_{t-1}, \mathbf{m} \mid \hat{r}_t, \hat{r}_{t-1}, \mathbf{z}_{1:t}, \mathbf{u}_{1:t}, \mathbf{w}) d\Delta \mathbf{x}_{t-1} d\mathbf{m} \end{aligned} \quad (6)$$

Under the assumption that the conditional distribution of  $\Delta \mathbf{x}_t$  can be approximated reasonably by a normal distribution, we can estimate the mean  $\Delta \bar{\mathbf{x}}_t$  of the marginal by maximum likelihood inference over the joint space of the relative poses  $\Delta \mathbf{x}_t$ , the map nodes  $\mathbf{m}$ , and the previous relative poses  $\Delta \mathbf{x}_{t-1}$ . The marginalization step is performed by simply extracting the values corresponding to  $\Delta \mathbf{x}_t$  from the joint mean vector. The covariance  $\Sigma_{t-1}$ , on the other hand, can be computed by linear error propagation through the optimization algorithm.

Note that the choice of reference frame (i.e., the chart) introduces a *map locality* due to the domain of the homeomorphism, implicitly discarding nodes and measurements of the graph not included in the domain.

Thus, dropping  $\hat{r}_t$  and  $\hat{r}_{t-1}$  for simplicity, the joint likelihood is proportional to

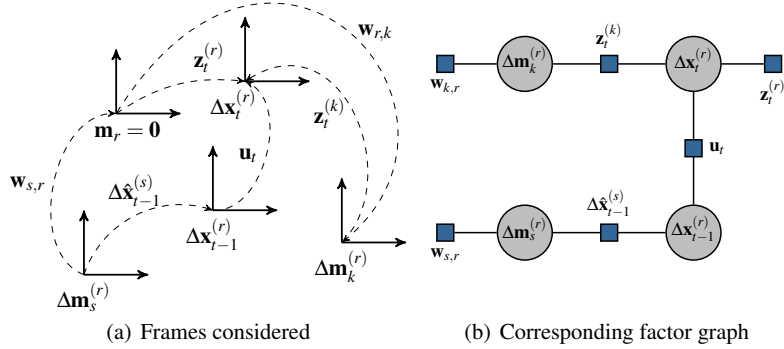
$$p(\mathbf{z}_t \mid \Delta \mathbf{x}_t, \mathbf{m}) p(\mathbf{w} \mid \mathbf{m}) p(\mathbf{u}_t \mid \Delta \mathbf{x}_t, \Delta \mathbf{x}_{t-1}) p(\Delta \hat{\mathbf{x}}_{t-1} \mid \mathbf{m}, \Delta \mathbf{x}_{t-1}), \quad (7)$$

where we assumed  $\mathbf{z}_t$ ,  $\mathbf{w}$ , and  $\mathbf{u}_t$  to be conditionally independent. Theoretically, we cannot assume independence between  $\mathbf{w}$  and  $\Delta \hat{\mathbf{x}}_{t-1}$ , since  $\Delta \hat{\mathbf{x}}_{t-1}$  was estimated at the previous time step by using  $\mathbf{w}$  as well. To avoid overconfidence due to this correlation, the resulting covariance estimate should be corrected. The correction factor can be computed following the approach of Mourikis and Roumeliotis [22]. We, however, experimentally found its effect to be negligible.

For ease of notation we will henceforth refer to  $\hat{r}_t$  as  $r$  and  $\hat{r}_{t-1}$  as  $s$ . Under the chart of  $r$ , the new reference node acts as the origin of its homeomorphic Euclidean space, with respect to which its neighborhood is expressed. We can thus consider a set of relative map pose coordinates  $\Delta \mathbf{m}_i^{(r)} \forall i \in V_t$ , which in the ideal case would be equivalent to  $\ominus \mathbf{m}_r \oplus \mathbf{m}_i$ .

To simplify the understanding of the conditional independence structure of the likelihood in (7), we provide in Fig. 2(a) the reference frames at work in the estimation problem. Here,  $k \in V_t$ , while the observations  $\mathbf{z}_t$  are assumed to provide SE( $n$ ) relative measurements, for example via ICP matching. In the general formulation, the latter is not required, and we commit to it only for the sake of clarity.

Under the assumption that  $\mathbf{z}_t$ ,  $\mathbf{u}_t$ , and  $\mathbf{w}$  are normally distributed, we can maximize the likelihood (7) by performing nonlinear least squares optimization on a re-



**Fig. 2** Frames and relative transformations involved in the small-scale estimation the robot’s relative pose (left) and the factor graph associated to the same problem (right).

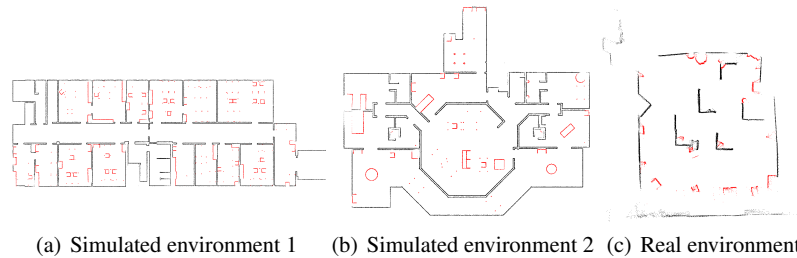
duced factor graph. Specifically, we construct a graph which has as vertices  $\Delta \mathbf{m}_s^{(r)}$ ,  $\Delta \mathbf{m}_k^{(r)} \forall k \in V_r$ ,  $\Delta \mathbf{x}_{t-1}^{(r)}$  and  $\Delta \mathbf{x}_t^{(r)}$ . The factors we introduce are respectively all the valid measurements  $\mathbf{z}_t^{(i)}$ , the odometry reading  $\mathbf{u}_t$ , the previous estimate  $\Delta \hat{\mathbf{x}}_{t-1}^{(s)}$ , and all  $\mathbf{w}_{i,j}$  connecting the map poses that have been introduced in the factor graph.

Note that there needs to exist a factor correlating  $\mathbf{m}_s^{(r)}$  to the remaining map vertices, as otherwise the previous estimate and the odometry reading would be neglected. If no such factor exists, it can be computed by linear error propagation.

Finally, we can compute the mean  $\Delta \bar{\mathbf{x}}_t^{(r)}$  of the next estimate by setting the current reference node to zero, conditioning the factor graph on it, and optimizing with respect to the remaining vertices. The resulting factor graph associated to the example in Fig. 2(a) is reported in Fig. 2(b). Since we are conditioning with respect to the current reference node, the factors that connected it have become priors on the other vertices. Note that in this formulation, there is no assumption on  $\mathbf{z}_t$  and it does not necessarily need to be an  $\text{SE}(n)$  measurement, however,  $\mathbf{z}_t$  should be such as to allow the overall problem to be non-singular and observable.

Having found the mean  $\Delta \bar{\mathbf{x}}_t^{(r)}$  by nonlinear least squares optimization, the estimation of the robot’s relative position to the remaining map poses is split over two methods. For the map poses  $\Delta \mathbf{m}_k^{(r)}$  that appeared in the factor graph optimization, we use the resulting mean estimate  $\Delta \bar{\mathbf{m}}_k^{(r)}$  and compute  $\Delta \bar{\mathbf{x}}_t^{(k)} = \ominus \Delta \bar{\mathbf{m}}_k^{(r)} \oplus \Delta \bar{\mathbf{x}}_t^{(r)}$ . For the remaining map poses we compute the relative estimate by chaining together  $\Delta \bar{\mathbf{x}}_t^{(r)}$  and all the transformations from the current reference node to the target map pose, over the minimum distance spanning tree of the full map (in this paper we used uniform weighting). Notice that with this formulation we explicitly compute only the uncertainty of  $\Delta \mathbf{x}_t^{(r)}$ , and we keep all other uncertainties as implicitly defined through linear error propagation.





**Fig. 3** Environments considered in the experiments in addition to the more common SLAM datasets. The objects marked in red do not appear in the unfurnished counterparts.

## 5 Experiments

We evaluated our relative topometric localization approach (RTL) on a 2D localization problem with range measurements. We considered virtual measurements computed by the Canonical Scan Matcher (CSM) with point-to-line metric [8] and its first order covariance approximation [7]. We further rely on  $g^2o$  [17] for the least squares optimization and on our MCL implementation for comparison [26].

We evaluated our approach in both simulation and real data. For a more realistic simulation, we also sourced laser data from real datasets, as performed by Olson [25]. For the simulation data, we employed the Gazebo simulator on an environment based on two real floor plans. For each environment we kept the unfurnished version, and created a further one with added obstacles and furniture. Fig. 3(a) and Fig. 3(b) show the two environments in question, with the additional obstacles and furniture marked in red. For each of the four maps we recorded one mapping run and ten localization runs connecting distant areas in the environment. For the pseudo-simulated data, we tested against the Intel Seattle, Intel Oregon, MIT CSAIL, and ACES datasets, which are publicly available. As ground truth, we used the aligned SLAM output and computed scans by casting rays on this aligned map. As with the simulated datasets, we recorded one mapping run and ten localization runs per configuration. In the above cases, we computed laser scans with  $360^\circ$  of field of vision and 360 rays. Each ray was corrupted with Gaussian noise with 1cm of standard deviation and rounded to the closest centimeter.

For the real robot experiment we created an environment in our lab, both with and without additional obstacles. We used a KUKA omnirob, equipped with two Sick S300 laser scanners with 541 beams and  $270^\circ$  of field of vision. The ground truth was approximated by using a motion capture system with 10 Raptor-E cameras recording at 300Hz. Again, we recorded one mapping run and 10 localization runs per configuration. To synchronize the data from the motion capture and the robot, we periodically stopped the robot and we only considered the poses in which the robot was stopped for both evaluation and mapping.

The motion capture may also introduce errors due to the not perfectly even floor, limited view-points of the cameras, and imperfect calibrations. Fig. 3(c) displays

the recorded laser scans aligned according to the motion capture as well as the additional obstacles that were introduced in red. Note that the map is imperfect due to the aforementioned issues, as such the quality of the evaluation for the real data is significantly lower than for the simulation datasets.

We compared our approach, RTL, with MCL as well as two additional benchmark approaches: a relative approach similar to RTL, where we assume the map is the result of an optimization process and is not subject to errors, (*LSLOC*) and MCL on the ground-truth map (*MCLGT*). Note that LSLOC, is equivalent in spirit to the approach of Konolige *et al.* [15] and Dayoub *et al.* [10]. For MCL and MCLGT we used 5000 particles, as measurement model we used likelihood fields saturated to 2m of maximum distance, and rendered the map at 1cm of resolution for maximum accuracy. We experimentally found that resolutions greater than 1cm did not improve the estimate. For each dataset, we use the same unoptimized pose graph as input to RTL and for computing the globally consistent map for MCL and LSLOC.

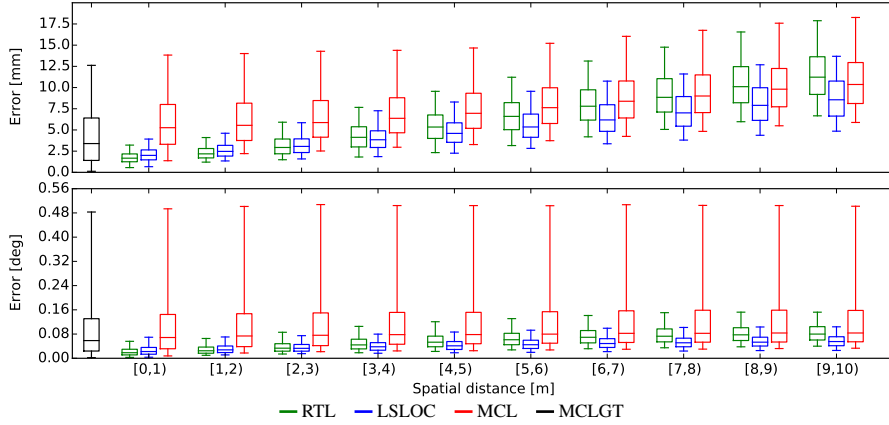
We consider three localization scenarios: *Standard localization*, where the given maps are globally consistent; *Localization with outliers*:, where the maps have different levels of inconsistency; and *Furnished vs. unfurnished localization*, where maps are globally consistent and we moved furniture between the mapping and localization phase. Since we are considering a tracking problem, we initialized the starting pose of all methods with the ground truth.

For the outlier case, we considered one simulated environment and the Intel Seattle dataset with two levels of outliers, respectively 5 and 20. We generate those outliers by associating the most likely matchings returned by FLIRT [31] that were at least at 3m of distance. A sample map with five outliers is reported in Fig. 1.

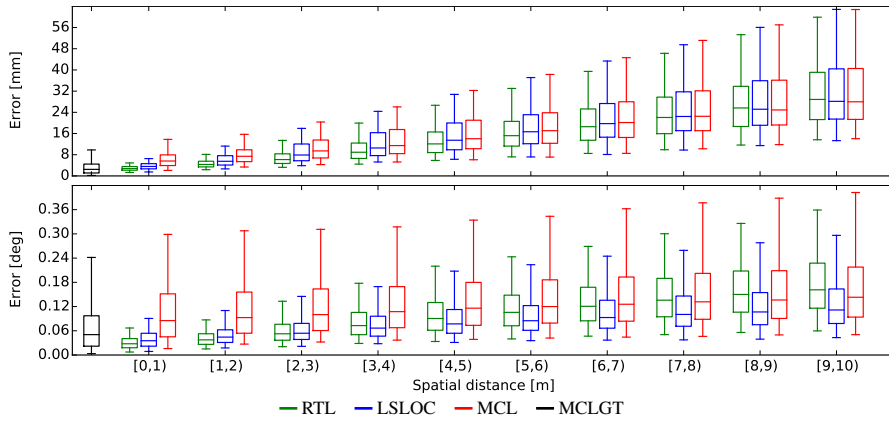
## 5.1 Evaluation criteria

We evaluate the localization accuracy at each time step  $t$  by measuring the discrepancy  $\epsilon_t^{(i)} = \ominus \check{\mathbf{x}}_t \oplus \check{\mathbf{m}}_i \oplus \Delta \mathbf{x}_t^{(i)}$  between the predicted relative displacements and the actual ones, for each reference frame  $i$ . Here  $\check{\mathbf{x}}_t$  and  $\check{\mathbf{m}}_i$  respectively refer to the ground-truth pose of the robot at time  $t$  and of the  $i$ -th element of the map.

We compute the error statistics according to the robot’s actual distance from the reference frame, in order to quantify the accuracy of all the methods at short distances, which are of interest to the robot, but also at progressively larger distances. We compute the shortest distances from the robot to the map poses by means of Dijkstra’s algorithm on the rasterized ground-truth map, with 1cm resolution. For each time step  $t$  we bin the absolute values of the errors  $\epsilon_t^{(i)}$  in terms of distance and use as statistics the 5th, 25th, 50th, 75th, and 95th percentile over the whole trajectory of all runs. When computing these statistics we do not take into account runs that diverged; we assume a method to diverge on any particular run if the average absolute error was greater than 1m in translation or  $60^\circ$  in rotation in the last 10% of the trajectory. Since MCLGT uses the ground-truth map for localization the statistics are independent of the distance, we thus report for it a single set of statistics.



**Fig. 4** Box plot of the errors in translation and rotation for the standard test scenario on the simulated datasets. The data was gathered over 40 localization runs.

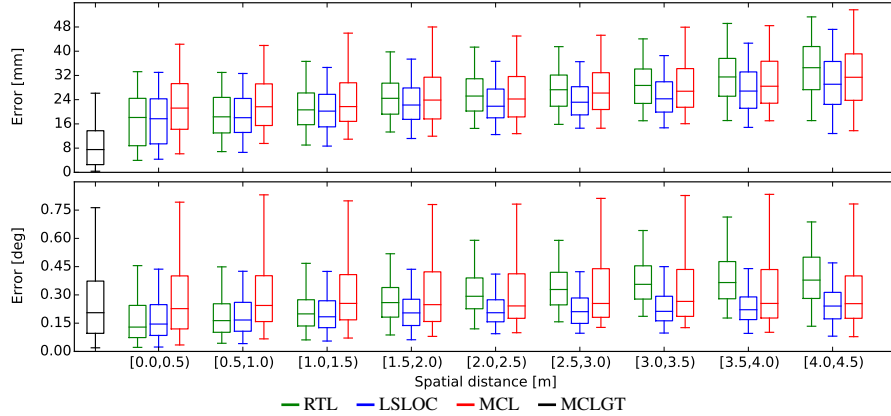


**Fig. 5** Box plot of the errors in translation and rotation for the standard test scenario on the pseudo-simulated datasets. The data was gathered over 40 localization runs.

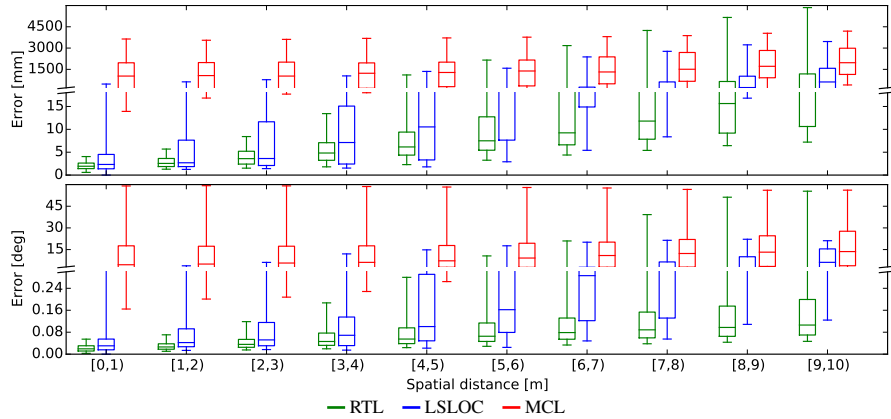
## 5.2 Accuracy results

We report in Fig. 4 and Fig. 5 the errors in terms of translation and rotation for the simulated and pseudo-simulated runs in the standard localization scenario. The box plots are binned at 1m of discretization, where the  $[a, b)$  values refer to all distances between  $a$  and  $b$ . The errors for the pseudo-simulated datasets are indeed larger than the simulated ones, which suggest an improved realism in the data sourcing.

As expected, the errors increase with respect to the distance for all methods, as even the most accurate SLAM map is subject to incremental error. At larger distances the error of RTL increases more as the local chart is increasingly inaccurate. This is to be expected and is the trade-off between local and global accuracy. RTL



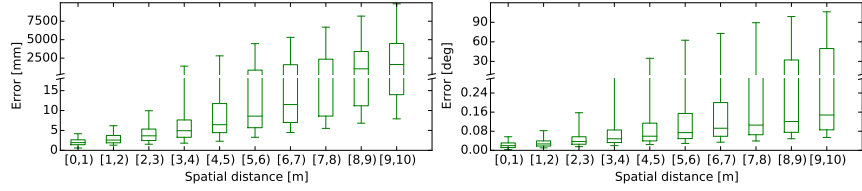
**Fig. 6** Box plot of the errors in translation and rotation for the standard test scenario on the real datasets. The data was gathered over 20 localization runs.



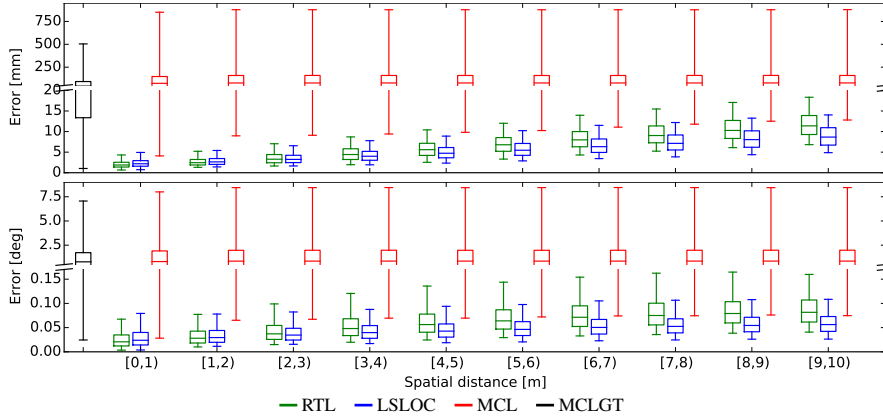
**Fig. 7** Box plot of the errors in translation and rotation for the test scenario on localization with 5 outliers. The data was gathered over 20 runs, with LSLOC having diverged 18 times and MCL 17 times. The error axis is broken in two scales, in order to display both small and large errors.

and LSLOC consistently achieve lower errors than MCL, particularly for orientation, and at short distances improve even on MCLGT. Nevertheless, given the scale of the graphs (millimeters and fractions of degrees) such differences are negligible.

Fig. 6 reports the accuracy results for the real datasets in the standard localization scenario. The errors are larger than both the pseudo-simulated and simulated datasets, partially due to the inaccuracy of the motion capture framework, the various unmodeled errors, and possibly due to the greater number of outliers in the laser range readings. Contrary to the previous tests, there is no clear preference to which method is most accurate. These results confirm the practical effectiveness of MCL when its assumptions are satisfied, with RTL showing competitive performance.



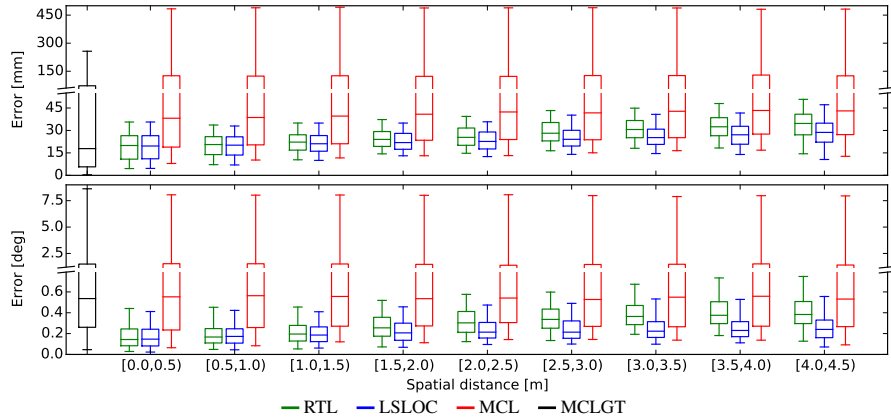
**Fig. 8** Box plot of the errors in translation and rotation of RTL for the test scenario on localization with 20 outliers. The data was gathered over 20 runs. RTL converged in all instances, the other methods diverged. The error axis is broken in two scales, to display both small and large errors.



**Fig. 9** Box plot of the errors in translation and rotation for the furnished vs. unfurnished test scenario on simulated datasets. The data was gathered over 40 runs. MCL diverged in 4 instances, MCLGT in 7. The error axis is broken in two scales, in order to display both small and large errors.

Fig. 7 shows the accuracy results for the localization scenario with 5 outliers. LSLOC and MCL respectively diverged 18 and 17 times out of 20 localization runs, and even when converged the methods exhibit significantly large errors in both position and orientation. This result showcases the brittleness of traditional approaches which rely on the assumption of a globally consistent map. On the contrary, RTL is virtually unaffected by the outliers up to approximately 4m of distance. At longer distances the relative position estimates may be obtained by chaining estimates through the outliers, which significantly degrades the accuracy. Similarly, Fig. 8 reports the accuracy results for the localization scenario with 20 outliers. LSLOC and MCL diverged in all runs, while RTL converged in all instances and shows similar performance. Although we have observed no such instances, in principle, RTL might be also trapped in a wrong reference frame and exhibit divergence, since we are committing on a single estimate for the reference frame. This problem can be counteracted by considering a distribution over the reference frames.

Fig. 9 and Fig. 10 report the errors in localization for the furnished vs. unfurnished test scenario, respectively on simulated and real datasets. For the simulated data MCL diverged in 4 instances, while MCLGT in 7.



**Fig. 10** Box plot of the errors in translation and rotation for the furnished vs. unfurnished test scenario on real datasets. The data was gathered over 20 runs. The error axis is broken in two scales, in order to display both small and large errors.

The MCL methods show significantly degraded accuracy when localizing on a map which does not reflect the structure of the perceived environment, as also shown in the literature [32]. We conjecture it is due to the fact that truncated likelihood fields and beam-based observation models are not robust enough to non-negligible discrepancies between the map and the observed environment. For future works it would be of interest to explore the conditions under which MCL results in degraded estimates and if it is possible to improve the observation model to account for them.

RTL and LSLOC, on the other hand, are virtually unaffected by the presence or absence of furniture. We believe this is due to the robustness of the ICP variant implemented in CSM and due to the fact that the localization is constrained to follow the connectivity of the graph and cannot “cross” a room wall, which can indeed happen with a standard variant of MCL.

### 5.3 Timing results

In our implementation, each localization step requires roughly 220ms of time on a multithreaded Intel® Core™ i7-3770K CPU clocked at 3.50GHz. By comparison, MCL with 5000 particles requires approximately 90ms. The current bottleneck is the number of ICP matchings executed per time step. This can be drastically reduced by introducing a better transition model  $p(r_t | r_{t-1})$ , which is currently merely uniform.

## 6 Conclusion

In this paper we introduced a novel localization approach, *relative topometric localization*, that relaxes the assumption of global consistency of the map used for localization. We represent the map as a pose graph endowed with sensory data for each node, which induces a manifold-like structure that, contrary to a globally consistent map, is not required to be fully embeddable in a Euclidean space. We reformulated the localization problem as estimating the topological node on the graph and a relative metric rigid body transformation. In doing so we obtained a method that is both robust to outliers in the map, takes into account the uncertainty in the mapping process, and that is independent of whether the graph has been optimized or not. We showed through extensive evaluation on both simulated and real data that our method not only significantly improves the state of the art in terms of localization accuracy when the global map consistency assumption fails, but also that for short relative distances it is equally accurate, or better, than MCL for the nominal cases.

**Acknowledgements** This work has been partially supported by the European Commission under the grant numbers ERC-AG-PE7-267686-LIFENAV, FP7-610603-EUROPA2, and H2020-645403-ROBDREAM.

## References

1. Agarwal, P., Tipaldi, G., Spinello, L., Stachniss, C., Burgard, W.: Robust map optimization using dynamic covariance scaling. In: Proc. of the IEEE Int. Conf. on Robotics and Automation (2013)
2. Andrew Howard, G.S.S., Matarić, M.J.: Multi-robot mapping using manifold representations. Proc. of the IEEE Special Issue on Multi-robot Systems **94**(9), 1360 – 1369 (2006)
3. Badino, H., Huber, D., Kanade, T.: Real-time topometric localization. In: Proc. of the IEEE Int. Conf. on Robotics and Automation (2012)
4. Bar-Shalom, Y., Li, X.R., Kirubarajan, T.: Estimation with applications to tracking and navigation: theory algorithms and software. John Wiley & Sons (2004)
5. Blackman, S.S.: Multiple hypothesis tracking for multiple target tracking. IEEE Aerospace and Electronic Systems Magazine **19**(1), 5–18 (2004)
6. Blanco, J.L., González-Jiménez, J., Fernandez-Madrigal, J.A.: Sparser relative bundle adjustment (SRBA): constant-time maintenance and local optimization of arbitrarily large maps. In: Proc. of the IEEE Int. Conf. on Robotics and Automation (2013)
7. Censi, A.: An accurate closed-form estimate of ICP's covariance. In: Proc. of the IEEE Int. Conf. on Robotics and Automation (2007)
8. Censi, A.: An ICP variant using a point-to-line metric. In: Proc. of the IEEE Int. Conf. on Robotics and Automation (2008)
9. Churchill, W., Newman, P.: Practice makes perfect? managing and leveraging visual experiences for lifelong navigation. In: Proc. of the IEEE Int. Conf. on Robotics and Automation (2012)
10. Dayoub, F., Morris, T., Upcroft, B., Corke, P.: Vision-only autonomous navigation using topometric maps. In: Proc. of the IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (2013)
11. Dellaert, F., Fox, D., Burgard, W., Thrun, S.: Monte Carlo localization for mobile robots. In: Proc. of the IEEE Int. Conf. on Robotics and Automation (1999)

12. Fox, D., Burgard, W., Thrun, S.: Markov localization for mobile robots in dynamic environments. *Journal on Artificial Intelligence Research* **11** (1999)
13. Grisetti, G., Kümmerle, R., Stachniss, C., Burgard, W.: A tutorial on graph-based SLAM. *IEEE Transactions on Intelligent Transportation Systems Magazine* **2**, 31–43 (2010)
14. Jensfelt, P., Austin, D., Wijk, O., Andersson, M.: Feature based condensation for mobile robot localization. In: *Proc. of the IEEE Int. Conf. on Robotics and Automation* (2000)
15. Konolige, K., Marder-Eppstein, E., Marthi, B.: Navigation in hybrid metric-topological maps. In: *Proc. of the IEEE Int. Conf. on Robotics and Automation* (2011)
16. Krüsi, P., Bücheler, B., Pomerleau, F., Schwesinger, U., Siegwart, R., Furgale, P.: Lighting-invariant adaptive route following using iterative closest point matching. *Journal on Field Robotics* (2014)
17. Kümmerle, R., Grisetti, G., Strasdat, H., Konolige, K., Burgard, W.: g2o: A general framework for graph optimization. In: *Proc. of the IEEE Int. Conf. on Robotics and Automation* (2011)
18. Latif, Y., Cadena, C., Neira, J.: Robust loop closing over time. *Proc. of Robotics: Science and Systems* (2012)
19. Leonard, J., Durrant-Whyte, H.: Mobile robot localization by tracking geometric beacons. *IEEE Transactions on Robotics* **7**(4), 376–382 (1991)
20. Mazuran, M., Tipaldi, G.D., Spinello, L., Burgard, W., Stachniss, C.: A Statistical Measure for Map Consistency in SLAM. In: *Proc. of the IEEE Int. Conf. on Robotics and Automation* (2014)
21. McManus, C., Furgale, P., Stenning, B., Barfoot, T.D.: Lighting-invariant visual teach and repeat using appearance-based lidar. *Journal on Field Robotics* (2013)
22. Mourikis, A.I., Roumeliotis, S.I.: On the treatment of relative-pose measurements for mobile robot localization. In: *Proc. of the IEEE Int. Conf. on Robotics and Automation* (2006)
23. Oh, S., Russell, S., Sastry, S.: Markov chain monte carlo data association for multi-target tracking. *IEEE Transactions on Automatic Control* **54**(3), 481–497 (2009)
24. Olson, E., Agarwal, P.: Inference on networks of mixtures for robust robot mapping. *Int. Journal of Robotics Research* (2013)
25. Olson, E.B.: Real-time correlative scan matching. In: *Proc. of the IEEE Int. Conf. on Robotics and Automation* (2009)
26. Röwekämper, J., Sprunk, C., Tipaldi, G.D., Stachniss, C., Pfaff, P., Burgard, W.: On the position accuracy of mobile robot localization based on particle filters combined with scan matching. In: *Proc. of the IEEE/RSJ Int. Conf. on Intelligent Robots and Systems* (2012)
27. Sibley, G., Mei, C., Reid, I., Newman, P.: Adaptive relative bundle adjustment. In: *Proc. of Robotics: Science and Systems* (2009)
28. Sprunk, C., Tipaldi, G.D., Cherubini, A., Burgard, W.: Lidar-based Teach-and-Repeat of Mobile Robot Trajectories. In: *Proc. of the IEEE/RSJ Int. Conf. on Intelligent Robots and Systems* (2013)
29. Strasdat, H., Davison, A.J., Montiel, J., Konolige, K.: Double window optimisation for constant time visual SLAM. In: *IEEE Int. Conf. on Computer Vision* (2011)
30. Sünderhauf, N., Protzel, P.: Switchable constraints for robust pose graph SLAM. In: *Proc. of the IEEE/RSJ Int. Conf. on Intelligent Robots and Systems* (2012)
31. Tipaldi, G.D., Arras, K.O.: FLIRT - interest regions for 2D range data. In: *Proc. of the IEEE Int. Conf. on Robotics and Automation* (2010)
32. Tipaldi, G.D., Meyer-Delius, D., Burgard, W.: Lifelong localization in changing environments. *Int. Journal of Robotics Research* **32**(14), 1662–1678 (2013)
33. Xu, D., Badino, H., Huber, D.: Topometric localization on a road network. In: *Proc. of the IEEE/RSJ Int. Conf. on Intelligent Robots and Systems* (2014)