Modeling movable objects improves localization in dynamic environments

Matti Pekkanen, Francesco Verdoja, and Ville Kyrki

***Abstract*— Most state-of-the-art robotic maps assume a static world; therefore, dynamic objects are filtered out of the measurements. However, this division ignores movable but non-moving, *i.e.*, semi-static objects, which are usually recorded in the map and treated as static objects, violating the static world assumption and causing errors in the localization. This paper presents a method for modeling moving and movable objects to match the map and measurements consistently. This reduces the error resulting from inconsistent categorization and treatment of non-static measurements. A semantic segmentation network is used to categorize the measurements into static and semi-static classes, and a background subtraction-based filtering method is used to remove dynamic measurements. Experimental comparison against a state-of-the-art baseline solution using real-world data from the Oxford Radar RobotCar data set shows that consistent assumptions over dynamics increase localization accuracy.**

I. Introduction

Most existing mapping methods assume that the mapped environment does not change until the map is used for localization. This is usually referred to as the *static world assumption*. The assumption is made for simplicity, even if it does not entirely hold. Violations of the assumption, however, may result in errors in the localization.

For example, the map might contain containers, scaffolding, or piled earth, which would be considered equally reliable landmarks compared to non-movable, *i.e.*, *static* objects such as buildings. If, during localization, another container was observed in a different pose than the container on the map that has since left, the potential incorrect match may cause a localization error. This phenomenon is illustrated in Figure 1.

To address this problem, many methods for removing moving, *i.e.*, *dynamic*, objects from the measurements have been proposed [1]–[4], and it continues to be the most common approach in the state-of-the-art localization and mapping methods. This dichotomy between moving and non-moving objects ignores objects that are *movable* while not currently moving.

In this work, we show a better way: by distinguishing between the properties of movability and motion, we can properly model the *dynamic classes*: dynamic, semi-static, and static. By consistently applying this distinction, we comply with not only the static world assumption but also all our assumptions over dynamics.

This work was supported by Business Finland, decision 9249/31/2021. We gratefully acknowledge the support of NVIDIA Corporation with the donation of the Titan Xp GPU used for this research.

M. Pekkanen, F. Verdoja and V. Kyrki are with School of Electrical Engineering, Aalto University, Espoo, Finland. {firstname.lastname}@aalto.fi

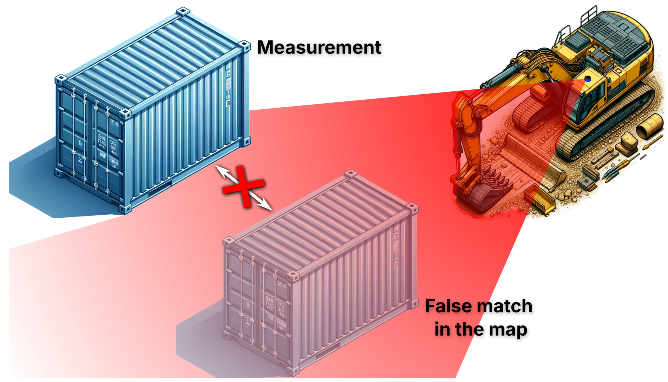

Fig. 1: Semi-static objects treated as static violate the static world assumption and cause mismatches between the map and the measurements. In this example, the map contains a container, which has since been moved away. When the robot returns, an observed container is offset from the one on the map. This offset causes matching errors, especially when the difference in poses is small or other features in the direction of the error are lacking or sparse.

We can partition the measurements into dynamic classes by using semantic segmentation of laser point clouds and background subtraction and clustering-based dynamic object filtering. Using these filters to be consistent in the assumptions over dynamics, we create a map containing only static measurements and four localization methods, each using measurements of different dynamic classes in the localization. Using real-world data from real traffic scenarios gathered over nine days; we show that localization under consistent assumptions over dynamics increases localization accuracy.

The main contributions of this paper are:

i) We propose a localization method using semantic segmentation and dynamic filtering to remove non-static measurements from the input measurements of the localization.
ii) We propose a mapping method using semantic segmentation to remove non-static measurements to produce a map compliant with the static world assumption.
iii) We show with an empirical study consisting of 112 localization experiments that the localization accuracy of the baseline method can be improved using the proposed mapping method to create a map consisting of only static measurements and the proposed localization method.

II. Related work

A. Filtering dynamic objects

The most commonly used map type in mobile robotics is the occupancy map. Occupancy maps incorporate the static world assumption, as they do not model the dynamic properties of the contents of the cells. Therefore, the sensor model cannot be adjusted to update the probabilities depending on the dynamic properties of the measurement and the affected grid cells.

To alleviate these problems, methods to filter dynamic objects from the measurements have been proposed [1], [2]. Even if dynamic objects are filtered from the measurements, unlike this work, none of these approaches distinguish between static and semi-static objects and subsequently leave the semi-static objects in the map.

B. Representation of semi-static objects

Several methods have been proposed to address the issue of semi-static objects being treated as static. Semi-static objects have been represented as separate temporary maps [5] with a given static map. While this is not done to create a consistent representation of the environment but rather to facilitate localization, this idea extends similar methods that jointly localize the robot and estimate the state of the environment, demonstrated multiple times with a door [6].

A step forward in representing the dynamic nature of the environment is to model it as an Hidden Markov Model (HMM) [7]. While an HMM models explicitly the belief of occupancy and the transition probabilities of the environment, which can be used to improve the localization accuracy, unlike this work, there is no distinction between dynamic or static cells.

Furthermore, the static world assumption is ingrained into the Markov assumptions of independence of odometry and observations. These assumptions have been relaxed by partitioning the localization experiment into internally Markovian episodes, but as a whole, they are not [8]. In this work, we aim instead to maintain a consistent environment representation.

While static objects are considered not movable, semi-static objects are likely to move during the lifetime of a map. Therefore, the probability of any object remaining stationary reduces over time. This degree of staticness can be modeled explicitly as the decaying probability of the persistence of a feature [9]. Features are more naturally linked to object instances that can be ascribed with a notion of staticness, whereas we directly model the dynamic properties of the entire spatial environment.

III. Problem statement

The generic localization problem is defined as finding the posterior distribution of $p(x_t|z_{0:t}, u_{0:t}, m_t)$, where $z_{0:t}$ is the sequence of sets of measurements $z_{0:t} = \{z_0, ..., z_t\}$, and m_t is the current state of the environment.

In localization, we commonly use a previously created map $m_{t_m} \approx m_t, t_m \ll t_0$. However, this approximation holds only for the static parts of the environment. To solve the posterior through Bayes' theorem, the problem is finding a model of the measurement likelihood $p(z_t|x_t, m_{t_m})$, which would take into account that semi-static and dynamic parts of the environment might have moved.

IV. Method

A. Definitions

To model the dynamics of objects, two properties of dynamics need to be considered: *movability* (whether an object can move) and *motion* (whether it is currently moving). The categorization of unmovable and movable objects depends on the context, *e.g.*, buildings can get demolished. However, we define unmovable objects as ones very unlikely to move during the lifetime of the map. We assume that the movability depends on the semantic label of the object.

We consider that objects can be separated into three dynamic classes: static $\mathcal{S}$, semi-static $\mathcal{E}$, and dynamic $\mathcal{D}$, defining the classes in terms of movability and motion as

- Static objects: objects that are unmovable.
- Semi-static objects: objects that are movable but not in motion.
- Dynamic objects: objects that are in motion.

We assume that movability is stationary over time; that is, objects that are unmovable cannot become movable, and vice versa. On the other hand, semi-static objects may start moving and become dynamic. Additionally, we assume that the dynamic properties are distinct and must be estimated independently. Therefore, if an object is not in motion, its movability cannot be inferred from that fact alone. These assumptions are consistent with the real properties of objects. Therefore, we call these *consistent assumptions over dynamics*.

B. Localization under consistent assumptions over dynamics

We propose to filter the measurements such, that the used measurements are consistent with the assumptions over dynamics. This enables likelihood estimation with consistent assumptions over dynamics with any measurement model and may be used in localization, mapping, or both.

The method consists of the following steps:

1) At time t, given the set of measurements z_t, a dynamic class d_z^i is estimated for each measurement $z_t^i \in z_t$ using a function $d(z_t^i)$.
2) Using the acquired dynamic classes, a subset of measurements $\tilde{z}_t \subseteq z_t$ is selected such that it consists of only the measurements belonging to a set of selected dynamic classes δ_z.

$$\begin{aligned}\tilde{z}_t &= \{z_t^i \in z_t : d(z_t^i) \in \delta_z\} \\ \delta_z &\subseteq \{\mathcal{S}, \mathcal{E}, \mathcal{D}\},\end{aligned}$$

When the method is applied in localization, using the acquired subset of measurements $\tilde{z}$, the original measurement model $p(\tilde{z}_t|x_t, \tilde{m})$ comprises the given set of assumptions over dynamics, defined by δ_z and δ_m. When the method is used when building a map, it yields a map $\tilde{m}$ that consists of only measurements of the selected dynamic classes δ_m. This formulation has the benefit of leaving the definitions of the function $d(z)$, the map m, and the model $p(z_t|x_t, m)$ open for various implementations while enforcing constraints over dynamics.

If the map is temporary, such as used in local collision avoidance, and not to be reused when returning to the same area, the assumptions differ, and therefore, semi-static objects may be recorded into the map as they are static during the lifetime of the map. However, in normal map-based localization, only static measurements should be recorded into the map, meaning using the selection $\delta_z = \delta_m = \{\mathcal{S}\}$, the localization is consistent over assumptions over dynamics.

V. Experiments

The two main questions we want to answer with the experiments are:

1) Does the localization accuracy increase when the dynamic properties of the environment are better represented in the content of the map or the measurements?
2) Does the localization accuracy decrease over time from map creation? Does this depend on the dynamic properties of the content of the map or the measurements?

To answer these questions, we performed a series of experiments. We tested the proposed mapping method against the baseline Normal Distributions Occupancy Map (NDT-OM). We used two sequences from the data set to create two maps, each with each method, for four maps. Four localization methods were assessed using seven sequences for each map, totaling 112 localization experiments.

A. Data set

The Oxford Radar RobotCar data set [10], [11] was used in the experiments, which consists of 32 sequences from seven different days over nine days where the same route is traversed. Nine sequences were selected from the data set: two for mapping and seven for localization, one from each day of the data set. Measurements from the left Velodyne 32E laser and odometry were used as inputs.

The semantic segmentation was obtained using RandLA-net [12], with a pre-trained model provided by the authors. The model was trained using Semantic KITTI data set [13] and therefore uses the labels from that set, which contain separate labels for corresponding semi-static and dynamic objects, such as a car and a moving car, but the network could not reliably detect dynamic objects.

B. Filtering

We use two filters to implement the function $d(z)$ for partitioning the measurements into the dynamic classes. A dynamic filter removes measurements originating from dynamic objects. The filter removes the ground plane and clusters the remaining points. The cluster centroids are stored and associated with the cluster centroids of the subsequent measurement. The estimated movement of the cluster centroids combined with the semantic labels was used to determine whether the cluster represents a dynamic or non-dynamic object.

TABLE I: The used localization methods

Name	**dynamic filter**	**semantic filter**	δ_z
baseline	-	-	$\{\mathcal{S}, \mathcal{E}, \mathcal{D}\}$
filtered	✓	-	$\{\mathcal{S}, \mathcal{E}\}$
static	-	✓	$\{\mathcal{S}\}$
combined	✓	✓	$\{\mathcal{S}\}$

Second, a semantic filter removes all measurements with non-static semantic labels. We consider labels 40–99 from Semantic KITTI as static.

C. Map creation

Two maps were created from two sequences, yielding a total of four maps. The first map is the baseline NDT-OM, created using all measurements. The map contains only static and semi-static objects, as NDT-OM removes the dynamic objects. The second map uses only static measurements using the semantic label filter. Both maps were created using NDT-OM fusion method [14] using ground truth poses of the data set.

D. Localization

To study the effect of the selection of δ_z, we pre-process the measurements using the filters presented in Section IV-B and localize using Normal Distributions Transform Monte-Carlo Localization (NDT-MCL) [15], creating four localization methods, presented in Table I: one with each filter, one without any filtering, and one with both filters.

E. Results

Several conclusions can be drawn from the results in terms of Absolute Trajectory Error (ATE), which are presented in Figure 2.

First, using the static map improves localization accuracy. With all methods except static localization, using the static map would be preferable as it reduces variance, improves the mean, or both. With static localization, the difference between the maps is negligible. This is likely due to the nature of NDT registration, where only matches between measurements and the map contribute to the cost. As there is no cost for unmatched cells, it matters less if the measurements are removed from the measurements or the map, as the reduction in error is similar. These results indicate that the static map increases performance in three of the four cases, and in one case, the performance stays the same. As the static map consists of only static measurements, this result concurs with the hypothesis that having consistent assumptions over dynamics increases localization accuracy.

Second, the filtering of the measurements during localization also improves localization accuracy, as dynamic objects may cause large-magnitude errors when incorrectly matched with the map. Compared to the baseline localization, the filtering methods have reduced mean, variance, or both, making them more desirable choices.

TABLE II: Mean ATEs

localization type	baseline map	static map
baseline	2.8391 m	2.1629 m
filtered	2.2503 m	1.9615 m
static	**2.1012 m**	2.1342 m
combined	2.2414 m	**1.8749 m**

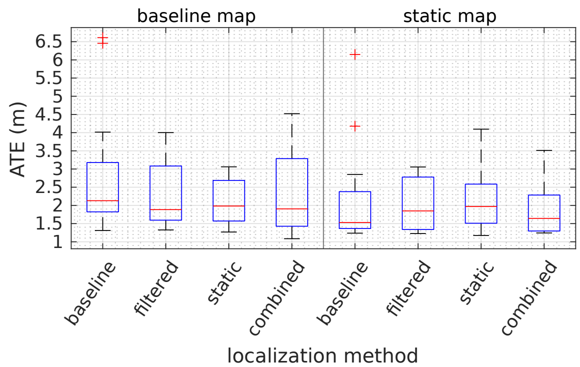

Fig. 2: The experiment results. In the figure the sample median is presented with a red line, and the blue box represents the range between 25th and 75th percentile, *i.e.*, the interquantile range. The black dashed line presents the interval between the minimum and the maximum samples. Values over 1.5 times the interquantile range are marked as outliers, and displayed with a red plus symbol.

Third, in terms of variance, static localization performs best. Whereas filtered localization can achieve very low errors, the variance is higher than static localization. While using more measurements is generally beneficial for localization accuracy, the incorrect matching of semi-static objects may cause errors. This makes the use of only static objects desirable, as they are the most reliable landmarks.

Given the two main hypotheses: (i) using only static measurements in the map and (ii) filtering the localization input are both beneficial for the localization accuracy, it should follow that the baseline localization with the baseline map should be the worst-performing combination, which can be seen from the results. As the baseline map holds semi-static measurements and the localization uses dynamic measurements, these can be incorrectly matched, reducing performance. Therefore, the localization accuracy is decreased by violating the consistent assumptions over dynamics. Conversely, when the static map and the combined method are used, the minimum ATE over all combinations is achieved.

VI. Conclusion

In this work, we argue that more realistic assumptions over dynamics are necessary. We showed that violating the static world assumption increases the localization error in terms of ATE due to the mismatch between the map and semi-static or dynamic measurements treated as static. While the data set in this work was gathered in a relatively static urban setting, the proposed methods would likely be even more useful in environments containing more semi-static objects, such as a construction site. The results pave the way for new interesting research topics. The use of more realistic models of dynamics could enable localization in more challenging environments where current methods fail. In this work, we studied only localization accuracy, but the proposed methods could improve performance in other critical application areas of mobile robotics, such as mapping and path planning.

VII. Acknowledgments

The image of the excavator and the container in Figure 1 was generated with ChatGPT.

References

[1] D. Fox, W. Burgard, S. Thrun, and A. B. Cremers, “Position Estimation for Mobile Robots in Dynamic Environments,” *AAAI/IAAI*, vol. 1998, p. 6, 1998.

[2] D. F. Wolf and G. S. Sukhatme, “Mobile Robot Simultaneous Localization and Mapping in Dynamic Environments,” *Autonomous Robots*, vol. 19, no. 1, pp. 53–65, July 2005.

[3] J. P. Saarinen, H. Andreasson, T. Stoyanov, and A. J. Lilienthal, “3D normal distributions transform occupancy maps: An efficient representation for mapping in dynamic environments,” *The International Journal of Robotics Research (IJRR)*, vol. 32, no. 14, pp. 1627–1644, Dec. 2013.

[4] X. Chen, A. Milioto, E. Palazzolo, P. Giguère, J. Behley, and C. Stachniss, “SuMa++: Efficient LiDAR-based Semantic SLAM,” in *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Nov. 2019, pp. 4530–4537, arXiv:2105.11320 [cs].

[5] D. Meyer-Delius, J. Hess, G. Grisetti, and W. Burgard, “Temporary maps for robust localization in semi-static environments,” in *2010 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. Taipei: IEEE, Oct. 2010, pp. 5750–5755.

[6] D. Schulz and W. Burgard, “Probabilistic state estimation of dynamic objects with a moving mobile robot,” *Robotics and Autonomous Systems*, vol. 34, no. 2-3, pp. 107–115, Feb. 2001.

[7] D. Meyer-Delius, M. Beinhofer, and W. Burgard, “Occupancy Grid Models for Robot Mapping in Changing Environments,” *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*, vol. 26, no. 1, pp. 2024–2030, Sept. 2021.

[8] J. Biswas and M. Veloso, “Episodic non-Markov localization: Reasoning about short-term and long-term features,” in *2014 IEEE International Conference on Robotics and Automation (ICRA)*. Hong Kong, China: IEEE, May 2014, pp. 3969–3974.

[9] D. M. Rosen, J. Mason, and J. J. Leonard, “Towards lifelong feature-based mapping in semi-static environments,” in *2016 IEEE International Conference on Robotics and Automation (ICRA)*. Stockholm, Sweden: IEEE, May 2016, pp. 1063–1070.

[10] W. Maddern, G. Pascoe, C. Linegar, and P. Newman, “1 Year, 1000km: The Oxford RobotCar Dataset,” *The International Journal of Robotics Research (IJRR)*, vol. 36, no. 1, pp. 3–15, 2017.

[11] D. Barnes, M. Gadd, P. Murcutt, P. Newman, and I. Posner, “The oxford radar robotcar dataset: A radar extension to the oxford robotcar dataset,” in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, Paris, 2020. [Online]. Available: https://arxiv.org/abs/1909.01300

[12] Q. Hu, B. Yang, L. Xie, S. Rosa, Y. Guo, Z. Wang, N. Trigoni, and A. Markham, “Randla-net: Efficient semantic segmentation of large-scale point clouds,” *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020.

[13] J. Behley, M. Garbade, A. Milioto, J. Quenzel, S. Behnke, C. Stachniss, and J. Gall, “SemanticKITTI: A Dataset for Semantic Scene Understanding of LiDAR Sequences,” in *Proc. of the IEEE/CVF International Conf. on Computer Vision (ICCV)*, 2019.

[14] T. Stoyanov, J. Saarinen, H. Andreasson, and A. J. Lilienthal, “Normal Distributions Transform Occupancy Map fusion: Simultaneous mapping and tracking in large scale dynamic environments,” in *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems*. Tokyo: IEEE, Nov. 2013, pp. 4702–4708.

[15] J. Saarinen, H. Andreasson, T. Stoyanov, and A. J. Lilienthal, “Normal distributions transform Monte-Carlo localization (NDT-MCL),” in *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. Tokyo: IEEE, Nov. 2013, pp. 382–389.