Precise Robot Localization in Architectural 3D Plans

Hermann Blum[1], Julian Stiefel[1,2], Cesar Cadena[1], Roland Siegwart[1] and Abel Gawel[1]

[1]Autonomous Systems Lab, ETH Zürich
[2]Multi Scale Robotics Lab, ETH Zürich

{blumh, jstiefel, cesarc, rsiegwart, gawela}@ethz.ch

Abstract -

This paper presents a localization system for mobile robots enabling precise localization in inaccurate building models. The approach leverages local referencing to counteract inherent deviations between as-planned and as-built data for locally accurate registration. We further fuse a novel camera-based robust outlier detector with LiDAR data to reject a wide range of outlier measurements from clutter, dynamic objects, and sensor failures. We evaluate the proposed approach on a mobile robot in a challenging real world site. In presence of clutter and model deviations, our system reduces the localization error by at least 32%.
A supplementary video summary can be accessed at `https://youtu.be/amqFPly8ZEQ`.

Keywords -

Construction Robotics, Localisation, On-Site Robotic Construction

1 Introduction

Assistive mobile robots in building construction enable both higher degrees of digitized processes, and reduce risk for human workers [1]. Construction robots therefore pose a high potential to transform the building construction process and are important facilitators of the ongoing effort for higher digitization. To this end, mobile robots perceive the environment with digital sensors, offering ease of relating information from digital building models to robot perception. Localizing robots in these building models with high accuracies then enables them to perform building tasks with respect to these data, or accurately track construction progress.

Conventional methods for robot localisation in construction rely on using external sensing, e.g., total stations, or augmentation of the environment, e.g., with artificial markers, to achieve high accuracies. However, these solutions require line-of-sight to a manually placed total station or marker and therefore depend on time-consuming manual preparation for every site, thus limiting the ease and autonomy of such systems. Furthermore, accurately localizing model information with mobile robots is not straightforward, due to the following challenges:

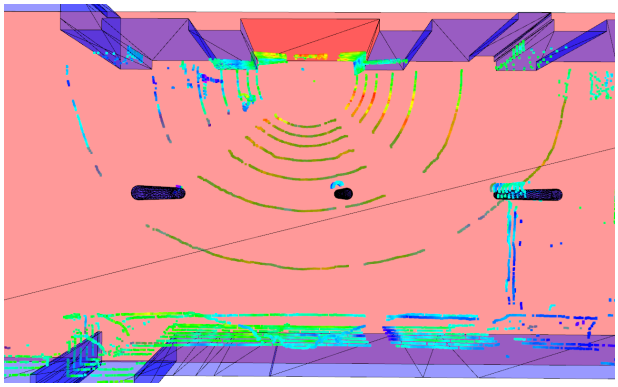


Figure 1: Our proposed method localizes a LiDAR scan against reference surfaces (red) in the mesh of a building model. The scan of the LiDAR is fused with semantic information and points are colored according to their high (yellow - red) or low (blue) probability of belonging to building structure. Note that points visible through the transparent ground plane may also appear red. The visible mismatches between the LiDAR scan and the columns are not a localization failure, but highlight the challenge of localizing within a built structure deviating from a planned model.

- Multi modality: Robotic sensor data is recorded with extereoceptive sensors, e.g. cameras and LiDARs, while building models are typically manually created 3D mesh models, rendering data integration and registration between these domains difficult.
- Deviations: As-built environments typically deviate from the as-planned state by up to several centimeters (cm), rendering global references insufficient for accurate task execution.
- Clutter: Real construction sites contain numerous temporary artefacts (e.g. equipment, scaffolding), as well as dynamic actors, such as human workers. This clutter is not modelled in the building models and therefore disturbs registration.

In this work, we propose a robotic localization system that addresses these challenges by combining locally referenced sensing with a learning-based sensor fusion approach for robust outlier rejection. The system therefore requires only on-board sensing without any artificial preparation of the site. We address the multi-modality

aspect by converting the mesh model into a sparse point-cloud which is easily referenced to sensor point-cloud data using standard scan-matching algorithms. Furthermore, we propose a task-based referencing solution, yielding locally accurate localization. Finally, we fuse a novel learning-based robust detector for outlier rejection on image data with LiDAR data, able to reject clutter that is outside the training distribution. Our system is tested on a mobile robot in realistic construction environments, showing a reduction of localization error of at least 30%. In summary, this paper presents the following contributions:

- Fusing range sensing data with learned visual outlier filters, producing semantically annotated point clouds that can be leveraged in 3D floorplan localization.
- A referencing system that disambiguates deviations between as-planned and as-built and improves localization accuracies for locally referenced building tasks.
- Evaluation of the proposed methods in real-world experiments on a robotic platform.

This paper is structured as follows: In Section 2, we present related works on robotic floorplan localization and semantic localization. In Section 3, we present our proposed 3D architectural floorplan method. We then report experimental results on a mobile robotic platform in Section 4, where these results are also discussed. Finally, we conclude our findings in Section 5.

2 Related Work

While registration of scans to building models is a well-studied problem in surveying [3], the application was studied less extensively in robotics. The important difference between these applications is that 3D scanners used in surveying remain stationary while scanning an environment (usually for at least some seconds), whereas registration on robotic systems is especially important while the robots move around. This *registration* or in the robotic sense *localisation*[1] is part of the state estimation that is required to accurately control the robot's movement. Robotic LiDAR sensors therefore are in general much faster 3D scanners that however also yield more sparse and imprecise scans and therefore pose unique challenges.

2.1 Robot Localization in Architectural Plans

With the raise of LiDAR sensors, different works studied 2D robot localization within floorplans [4, 5]. However, to execute construction tasks, 2D localization is insufficient. Walls might be tilted or floors elevated such that water can run off. At higher levels of accuracy, the flat and rectangular world assumption therefore is no longer valid. [6] therefore studied the extension of conventional methods to 3D, but rely on a special measurement system to localize their robot's endeffector with respect to local reference walls. In this work, we study localization methods that do not require manipulators mounted on the robot. Bosché [7] shows ICP-based localization of 3D scan data within building mesh models under the assumption of negligible deviations between model and reality.

In case that LiDAR sensors are not available, research has made progress to extract floorplan-like information out of camera images [8]. Boniardi et al. [9] studied how such information can be used to localize a robot in 2D within a given floorplan. Mendez at al. [10] localize an RGB camera in floorplans based on semantic information and can show that their method does not benefit from range information.

2.2 Semantically Enriched Localization with LiDARs

Global localization techniques can be robustified by using semantic cues [11]. However, the considered semantics only considered distinct classes as information cue, while our approach is more fine-grained using semantic information for filtering and weighting measurements. Closest to our work is [12]. The authors propose to use a semantic segmentation network on LiDAR data and a semantic consistency term in a surfel-based map representation to filter dynamic objects over multiple observations. Furthermore, the work proposes to weigh associations of an ICP registration using semantic label classification scores. Our work differs in the used network that does not require pre-knowledge about semantic classes, and can generalise to reliably detecting outliers outside its training distribution.

2.3 LiDAR Registration

Among the first papers, [13] describe the use of ICP for registration of 3D shapes, i.e., an iterative error minimization over 3D point-correspondences. A multitude of variants were proposed since, including more generalized formulations also including probabilistic measurements [14], and variants that skip the re-association step [15]. A comprehensive overview is given in [16]. Further notable registration algorithms rely on feature extraction [17], or representing data as a probability density [18]. [19] demonstrate ICP for localization with respect to a known object in the environment using a LiDAR scanner, achieving sub-centimeter accuracies. [20] achieve sub-cm accuracies using 2D LiDAR localization in a reference scan of the environment. However, the authors also introduce clutter and dynamics to the scenes which heavily degraded

[1]For the purpose of this work, registration and localisation are used synonymous.

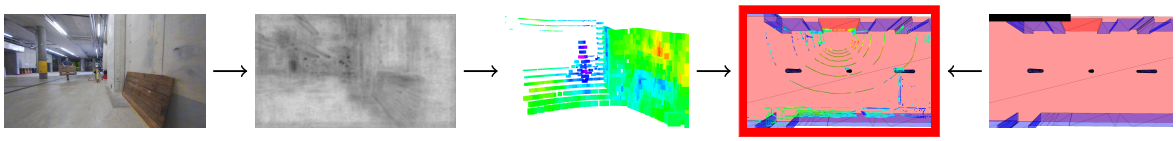


Figure 2: Overview of the proposed method: RGB images are segmented into foreground and background using a robust segmentation network [2]. These pixelwise scores are then propapgated and fused with LiDAR scans. The classifications are then used as weights in the floorplan localization which performs a weighted selective localization in a 3D floorplan given local task references.

the achievable accuracies. Without modifying the principal ICP solution, our method proposes to use only partial data, i.e., local references for scan registration. Furthermore, we explicitly consider outliers.

More robust pointcloud registration such as TEASER [21] have been proposed, but rely on 3D Keypoint descriptors, which perform poorly in manhatten-world building plans.

3 Method

An overview about the proposed method can be found in Figure 2. In the following, we first describe the input processing pipeline for LiDAR scans and define the general problem of registering and aligning scans to building models. Subsequently, we describe how further information on the task references can be used to define the alignment more accurately. The whole method assumes a good initial guess for the robot pose to coarsely align the building plan with the initial pose of the robot. This initialization can be provided manually or with a global localization method.

3.1 Semantic Filtering

Classic semantic segmentation of images or pointclouds relies on large datasets to train networks to reasonable performace. Unfortunately, there are no such labelled datasets available for construction environments. Moreover, construction environments are highly dynamic with a lot of potential object classes appearing, whereas we are only interested in background-foreground segmentation, i.e., separating building surfaces from any other scene contents. Due to the limits of available datasets and the potential problems with training on a fixed set of classes, we use an alternative segmentation method described in [2]. The density estimation network from Marchal et al. is trained on the NYU Indoor Room Dataset [22], but able to generalize better than classical methods to new environments. In particular, we use their best performing model, which is a regression over density estimation at multiple layers of the feature extractor. Instead of a binary segmentation, the method returns a per-pixel score that is higher if the pixel belongs to background structure of a building. We will hereafter refer to this score as density value d.

The proposed localization system relies on a multi-camera plus LiDAR setup with known intrinsic, extrinsic and timestamp calibrations. Whenever a new LiDAR scan arrives, we rectify the corresponding set of images $I = \{I_0, I_1, ...\}$ and process them in the density estimation network [2].
From the scanned LiDAR points $P = \{p_1, p_2, ...\}$, we project each point onto each image plane. For those points p_i within the field of view of a camera, we assign the density value d_i to the point. We reject all points that cannot be projected onto any of the images, but select camera lenses such that the amount of such points is negligibly small.

To localize within the building model, the scan P is aligned to the model with point-to-plane ICP:

$$T_{\text{icp}} = ICP(P, S) = \arg\min_{T} \sum_{i} w_i \, c\,[(p_i - m_i) \cdot \boldsymbol{n}(m_i)]$$

where $c()$ is a cost function, m_i is a matched point of p_i in the building model S and $\boldsymbol{n}(m_i)$ is the surface normal at that point.

For the weight w, we propose two variants:

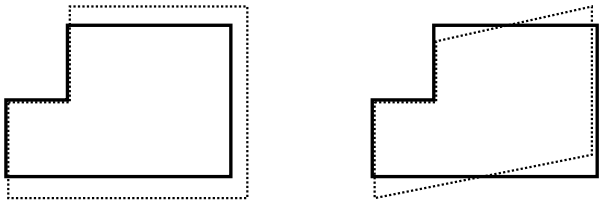
$$w_i = \begin{cases} 0 & d_i < \delta \\ 1 & d_i \geq \delta \end{cases} \tag{1}$$

$$w_i = \max(0, a\, d_i - \delta') \tag{2}$$

Where a is a normalization factor such that $\max_i w_i = 1$. Equation (1) corresponds to a binary segmentation of the scan where only those points which belong to background structure are considered in the ICP problem (*masked clutter*) and equation (2) assigns higher weights to points that are more likely to belong to background structure (*weighted clutter*). The non-binary weighting does not rely on perfect predictions from the density estimation network and instead uses the prediction confidence in the ICP optimization.

3.2 (Selective) Localization to References

We define the mesh of the building model as a collection of closed surfaces $S = \{s_0, s_1, ...\}$. In general, we can then



(a) ambiguours alignment (b) high-error alignment

Figure 3: Illustrations for deviations between as-planned (solid) building model and as-built scan from the robot (dotted). In (a), the alignment of the horizontal walls is ambiguous and traditional ICP will align that wall with more points in the scan. In (b), any alignment will yield high errors and the solution in general depends on the settings for outlier filtering.

localize the robot with $ICP(S, P)$. However, building models are usually far from perfect maps of the actual environment, where during the construction phase walls might be missing, temporary structure such as scaffolding is not mapped and the relative positions of the final walls can deviate from the building model. Some of the typical deviations are illustrated in Figure 3.

To enable the robot to localize precisely in a building that deviates from the map, we define a subset of walls or surfaces $R \subset S$ with $\exists\ r_1, r_2, r_3 : r_1 \nparallel r_2, r_2 \nparallel r_3, r_1 \nparallel r_3, r_1, r_2, r_3 \in R$. The reference surfaces therefore define a (minimally) sufficient alignment problem for the robot to localize itself, but exclude all surfaces in S that are not relevant to a local task and would in case of deviations disturb the alignment. This is further motivated by most construction tasks being expressed in local reference frames. Within the case of approximately rectangular building structures that we study in our experiments, R is often the set of surfaces defining a corner in a room. Such references are part of the task definition and can e.g. be synchronised online with robot [23].

Initial experiments with our methods showed that in cases where the surfaces in R are only measured with a few points of the LiDAR scan P, the ICP alignment can fail with huge errors. We therefore propose the following localisation procedure:

1. Find the alignment $T^{(t)}_{\text{full mesh}} = ICP(S, P|T^{(t-1)})$
2. Refine the alignment to the references $T^{(t)}_{\text{references}} = ICP(R, P|T^{(t)}_{\text{full mesh}})$
3. Reject $T^{(t)}_{\text{references}}$ if $\left|T^{(t)}_{\text{references}} - T^{(t)}_{\text{full mesh}}\right|$ is too large.

4 Experiments

4.1 Experimental Setup

We conduct our experiments on the *supermegabot*[2] platform with a LiDAR and three cameras. The robot is further equipped with an IMU for smooth state estimation. The exact sensor configuration is shown in Figure 4.

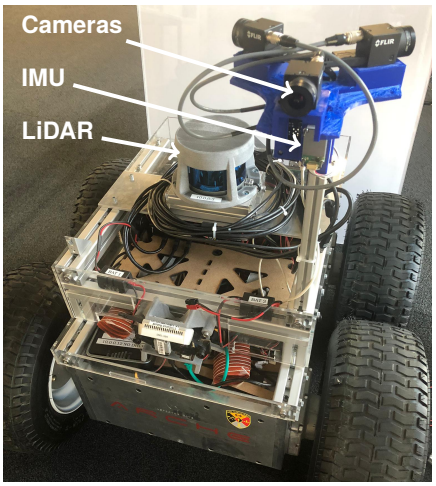


Figure 4: Sensor setup on the robot.

We calibrate all cameras individually using Kalibr[3] [24, 25] to obtain intrinsics and extrinsics with respect to the IMU. We then record a joined motion of LiDAR and cameras to find their respective transformations using visual-intertial odometry [26] with one of the cameras and optimizing the alignment of LiDAR scans with respect to that trajectory[4]. Time-synchonization is achieved with the VersaVIS [27] camera trigger board that synchronizes time between the host, all cameras and the IMU, while negligible time-offset is assumed between the host and the LiDAR.[5]

We supply the robot with a 3D mesh of the building that is generated from the available 2D floorplan. Because there are no information of the floor or the ceiling structure available, we add a planar floor to the mesh and give all walls the same height. Reference surfaces are specified as a subset of the same mesh.

To measure ground-truth, we attach a prism to the robot and measure the position with a total station referenced to the origin of the building plan that the robot is using as a map. We calibrate the prism position with respect to the robot frame by aligning trajectories of the robot moving around. In case that as-built reference walls deviate from the building plan in absolute coordinates, we measure this deviation and correct the ground truth position accordingly.

4.2 The Effect of Clutter

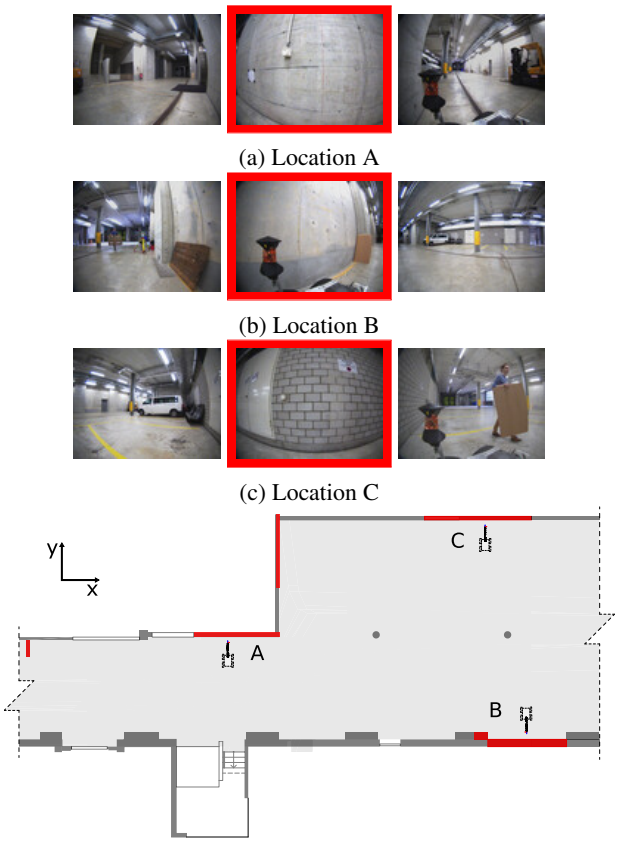
To verify the influence of clutter objects and moving workers on the general localization performance of

[2] github.com/ethz-asl/eth-supermegabot

[3] github.com/ethz-asl/kalibr

[4] github.com/ethz-asl/lidar-align

[5] While complicated, this procedure is necessary such that the image segmentation is correctly propagated into the LiDAR scan.



(a) Location A

(b) Location B

(c) Location C

(d) Site Overview

Figure 5: Selected locations and reference walls (marked red in d) for our experiments. All locations also have the floor as a reference surface. The red border around the images corresponds to the forward-facing camera.

a robotic system, we compare localisation at the same positions with and without clutter in Table 1.

We conduct experiments with a stationary robot while a construction worker is moving around the robot and clutter objects such as wooden boards and equipment boxes are placed to partially obstruct measurements to reference walls. For each location, we analyze a sequence of around 1 min. or 300 LiDAR scans.

The investigation shows that on-site clutter negatively affects the localisation accuracy and therefore confirms our initial hypothesis.

4.3 The Effect of As-Built Deviations

To simulate a severe deviation between as-planned and as-built, we move the upper and lower structure in Figure 5 apart by 0.3m in the mesh supplied to the robot. While the initial building mesh without this added deviation is already not an accurate representation of the environment, this further augmentation is used to compare localisation accuracy accross higher and lower deviations.

clutter	Loc. A	Loc. B
no	157	74
yes	218	83

Table 1: Influence of clutter on the localization accuracy (RMSE in mm). For this comparison, the robot localizes the full LiDAR scan in the full building model. At location C, the clutter was fixed to the environment and therefore a comparison without clutter could not be done.

deviations	Loc. A	Loc. B	Loc. C
lower	218	83	87
higher	390	222	76

Table 2: Influence of deviations between as-planned and as-built on the localization accuracy (RMSE in mm). The robot localizes the full LiDAR scan in the full building model, in a cluttered environment.

The localisation results of this comparison are listed in Table 2. In two locations, the additional deviation severely worsens the localisation results, while it has little influence on location C. It is expected that particular deviations affect localisation at some locations more than at others. In general, the results confirm the suspected effect of deviations between as-planned and as-built on the localisation accuracy.

4.4 Accuracy of the Proposed System

Given the negative effects of deviations and clutter, we now introduce our robotic system to a cluttered environment with the stated deviations between as-planned and as-built. We then analyse the localisation accuracy to investigate to what extend our introduced mitigation measures can compensate for the added difficulties.

We compare all combinations of methods described in section 3 to a baseline of ICP alignment between the full building model and the full LiDAR scan. We measure repeatability by the trace and maximum eigenvalue of the covariance matrices for position and rotation. For accuracy, we compare the estimated position of the robot against the total-station tracked prism. This measurement therefore relies on the estimation of the full robot pose. The results are given in tables 3, 4 and 5 for the three different test locations respectively. Additionally, we report the fraction of LiDAR scans that could not be localized either because ICP did not converge or the solution was rejected (see section 3.2). To account for non-determinisms in the localization pipeline, all values are averages over three executions on the same input data.

From the results we see that in general, selective localization against reference walls constraints the localization very well in two dimensions. The trace and maximum eigenvalue for localisation against reference surfaces are very close, whereas there is uncertainty in more than one

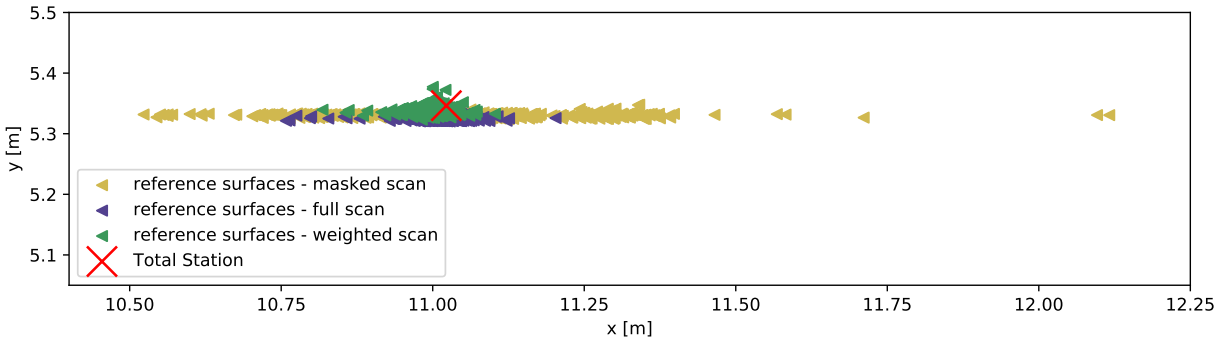


Figure 6: Topview on position C of the distribution of localization outputs for the stationary robot. All points are transformed to the estimated position of the tracking prism. The ground-truth location is given by the output of the total station for the tracked prism.

ICP Input		Pos. Repeatability		Rot. Repeatabilty		Accuracy	Failure Rate
map	scan	max eigenval.	trace	max eigenval.	trace	rmse [mm]	[%]
full mesh	full	22.3	25.9	**2.7**	**8.0**	390	0.0
full mesh	masked clutter	**1.5**	**2.5**	2.9	8.6	290	4.9
full mesh	weighted clutter	10.1	12.7	3.0	8.9	354	8.2
references	full	121.1	127.2	5.3	15.9	452	50.3
references	masked clutter	31.9	32.2	5.1	15.2	**232**	44.1
references	weighted clutter	330.5	338.6	14.1	42.4	627	76.5

Table 3: Stationary Localization Study Position A

ICP Input		Pos. Repeatability		Rot. Repeatabilty		Accuracy	Failure Rate
map	scan	max eigenval.	trace	max eigenval.	trace	rmse [mm]	[%]
full mesh	full	3.3	5.9	**3.2**	**9.6**	222	0.0
full mesh	masked clutter	**1.4**	**2.3**	3.4	10.3	342	5.3
full mesh	weighted clutter	3.8	5.6	3.6	10.7	230	8.1
references	full	7.6	8.0	3.4	10.3	328	4.8
references	masked clutter	2.6	2.7	4.2	12.6	**68**	21.3
references	weighted clutter	17.1	18.4	7.0	21.0	244	52.6

Table 4: Stationary Localization Study Position B

ICP Input		Pos. Repeatability		Rot. Repeatabilty		Accuracy	Failure Rate
map	scan	max eigenval.	trace	max eigenval.	trace	rmse [mm]	[%]
full mesh	full	1.8	2.9	**4.3**	**12.9**	76	0.0
full mesh	masked clutter	14.8	17.1	4.4	13.3	153	3.7
full mesh	weighted clutter	4.0	5.2	4.6	13.7	103	6.4
references	full	**0.9**	**1.0**	**4.3**	13.0	65	0.3
references	masked clutter	28.7	28.9	5.9	17.6	154	25.6
references	weighted clutter	1.0	1.5	4.8	14.4	**52**	9.8

Table 5: Stationary Localization Study Position C

direction for registration against the full building model. The same effect is illustrated in figure 6. This raises the question why the localization shows such a high variance in lateral direction in all three locations. A more detailed evaluation of the test site shown in figure 5 reveals that most of the walls in the building model are parallel to the x axis of the model and much fewer structures are parallel to the y axis. For all 3 locations, however, the lateral localization depends on the alignment of surfaces parallel to the y axis. The biggest available structure in this direction, which is also the reference wall for location C, is obstructed by a van parked in front of the wall. Therefore, for all three locations, the observable part of the lateral reference structure is small compared to both other directions and a higher localization uncertainty in x direction had to be expected.

Our experiments further show that semantic filtering is more important when localising against reference surfaces than when using the full mesh. Intuitively, localising only against references is more vulnerable to obstructions of these surfaces. While localisation in the full mesh is also worse in presence of clutter, as shown in table 1, our experiments show in two of three locations that without semantic filtering of the clutter, this impact is even more severe when localising against reference surfaces.

In all locations, the highest accuracy is achieved with a combination of semantic filtering of the LiDAR scan and selective localization to reference surfaces. The accuracy gain ranges from 32% in position C to 69% in position B. However, we could not find a single combination of methods that always worked. We found the performance of the semantic filtering method to be highly dependent on the location and the respective performance of the density estimation network. As we already describe in section 3.1, semantic classification in construction environments is difficult due to the lack of labelled data. The density estimation network from [2] showed reasonable performance for our test location, but still partially filtered out background structure that was too different from the training domain. In particular in location C, binary filtering of the pointcloud removed nearly all points on the lateral reference wall, resulting in high uncertainty and failure rate. We conclude that for precise localization of robots on construction sites, outlier filtering and selection of high quality measurements are key aspects. While learning-based solutions show promising characteristics in this regard, they require more domain-specific training data to further boost their performance.

5 Conclusion

In this work, we present a mobile on-board localization system for construction robots. In our experiments, we show that building deviations and clutter deteriorate accuracy of traditional registration methods, and find that a combination of semantic filtering of LiDAR scans and selective localization to reference walls yields essential accuracy gains. Our findings show a need for semantic datasets closer to the construction domain and a need for more research to further close the accuracy gap between external reference systems and on-board sensing.

Acknowledgement

This work was partially supported by the Swiss National Science Foundation (SNF), within the National Centre of Competence in Research on Digital Fabrication and by the HILTI group.

We would like to thank Selen Ercan for creating the 3D building models for our experiments. We would also like to thank Francesco Sarno for working with us on initial segmentation experiments.

References

[1] Hadi Ardiny, Stefan Witwicki, and Francesco Mondada. Are autonomous mobile robots able to take over construction? a review. *International Journal of Robotics, Theory and Applications*, 4(3):10–21, 2015.

[2] Nicolas Marchal, Charlotte Moraldo, Hermann Blum, Roland Siegwart, Cesar Cadena, and Abel Gawel. Learning densities in feature space for reliable segmentation of indoor scenes. *IEEE Robotics and Automation Letters*, 5(2):1032–1038, 2020.

[3] Gary KL Tam, Zhi-Quan Cheng, Yu-Kun Lai, Frank C Langbein, Yonghuai Liu, David Marshall, Ralph R Martin, Xian-Fang Sun, and Paul L Rosin. Registration of 3d point clouds and meshes: A survey from rigid to nonrigid. *IEEE transactions on visualization and computer graphics*, 2012.

[4] Federico Boniardi, Tim Caselitz, Rainer Kummerle, and Wolfram Burgard. Robust LiDAR-based localization in architectural floor plans. In *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2017. doi:10.1109/IROS.2017.8206168.

[5] W Hess, D Kohler, H Rapp, and D Andor. Real-time loop closure in 2D LIDAR SLAM. In *2016 IEEE International Conference on Robotics and Automation (ICRA)*, 2016. doi:10.1109/ICRA.2016.7487258.

[6] Abel Gawel, Hermann Blum, Johannes Pankert, Koen Krämer, Luca Bartolomei, Selen Ercan, Farbod Farshidian, Margarita Chli, Fabio Gramazio, Roland Siegwart, Marco Hutter, and Timothy Sandy.

A Fully-Integrated sensing and control system for High-Accuracy mobile robotic building construction. In *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*.

[7] Frédéric Bosché. Automated recognition of 3d cad model objects in laser scans and calculation of as-built dimensions for dimensional compliance control in construction. *Advanced engineering informatics*, 24(1):107–118, 2010.

[8] Ameya Phalak, Zhao Chen, Darvin Yi, Khushi Gupta, Vijay Badrinarayanan, and Andrew Rabinovich. DeepPerimeter: Indoor boundary estimation from posed monocular sequences. April 2019.

[9] Federico Boniardi, Abhinav Valada, Rohit Mohan, Tim Caselitz, and Wolfram Burgard. Robot localization in floor plans using a room layout edge extraction network. *arXiv preprint arXiv:1903.01804*, 2019.

[10] Oscar Mendez, Simon Hadfield, Nicolas Pugeault, and Richard Bowden. SeDAR - semantic detection and ranging: Humans can localise without LiDAR, can robots? September 2017.

[11] Renaud Dubé, Andrei Cramariuc, Daniel Dugas, Juan Nieto, Roland Siegwart, and Cesar Cadena. SegMap: 3d segment mapping using data-driven descriptors. In *Robotics: Science and Systems (RSS)*, 2018.

[12] Xieyuanli Chen, Andres Milioto, Emanuele Palazzolo, Philippe Giguère, Jens Behley, and Cyrill Stachniss. Suma++: Efficient lidar-based semantic slam. In *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 4530–4537. IEEE, 2019.

[13] Paul J Besl and Neil D McKay. Method for registration of 3-d shapes. In *Sensor fusion IV: control paradigms and data structures*, volume 1611, pages 586–606. International Society for Optics and Photonics, 1992.

[14] Aleksandr Segal, Dirk Haehnel, and Sebastian Thrun. Generalized-icp. In *Robotics: science and systems*, volume 2, page 435. Seattle, WA, 2009.

[15] Qian-Yi Zhou, Jaesik Park, and Vladlen Koltun. Fast global registration. In *European Conference on Computer Vision*, pages 766–782. Springer, 2016.

[16] François Pomerleau, Francis Colas, Roland Siegwart, and Stéphane Magnenat. Comparing icp variants on real-world data sets. *Autonomous Robots*, 34 (3):133–148, 2013.

[17] Ji Zhang and Sanjiv Singh. Loam: Lidar odometry and mapping in real-time. In *Robotics: Science and Systems*, volume 2, 2014.

[18] Peter Biber and Wolfgang Straßer. The normal distributions transform: A new approach to laser scan matching. In *IROS*, 2003.

[19] Timothy Sandy, Markus Giftthaler, Kathrin Dörfler, Matthias Kohler, and Jonas Buchli. Autonomous repositioning and localization of an in situ fabricator. In *2016 IEEE International Conference on Robotics and Automation (ICRA)*, 2016.

[20] Jörg Röwekämper, Christoph Sprunk, Gian Diego Tipaldi, Cyrill Stachniss, Patrick Pfaff, and Wolfram Burgard. On the position accuracy of mobile robot localization based on particle filters combined with scan matching. In *IROS*, 2012.

[21] Heng Yang, Jingnan Shi, and Luca Carlone. TEASER: Fast and certifiable point cloud registration. January 2020.

[22] Nathan Silberman, Derek Hoiem, Pushmeet Kohli, and Rob Fergus. Indoor segmentation and support inference from rgbd images. In *European conference on computer vision*, pages 746–760. Springer, 2012.

[23] Selen Ercan, Hermann Blum, Abel Gawel, Roland Siegwart, Fabio Gramazio, and Matthias Kohler. Online synchronization of building model for on-site mobile robotic construction. In *ISARC*, 2020.

[24] P Furgale, J Rehder, and R Siegwart. Unified temporal and spatial calibration for multi-sensor systems. In *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 1280–1286, November 2013. doi:10.1109/IROS.2013.6696514.

[25] J Maye, P Furgale, and R Siegwart. Self-supervised calibration for robotic systems. In *2013 IEEE Intelligent Vehicles Symposium (IV)*, pages 473–480, June 2013. doi:10.1109/IVS.2013.6629513.

[26] T Schneider, M Dymczyk, M Fehr, K Egger, S Lynen, I Gilitschenski, and R Siegwart. Maplab: An open framework for research in Visual-Inertial mapping and localization. *IEEE Robotics and Automation Letters*, 3(3):1418–1425, July 2018. ISSN 2377-3766. doi:10.1109/LRA.2018.2800113.

[27] Florian Tschopp, Michael Riner, Marius Fehr, Lukas Bernreiter, Fadri Furrer, Tonci Novkovic, Andreas Pfrunder, Cesar Cadena, Roland Siegwart, and Juan Nieto. Versavis: An open versatile multi-camera visual-inertial sensor suite. *arXiv preprint arXiv:1912.02469*, 2019.