

Real-time Multi-Adaptive-Resolution-Surfel 6D LiDAR Odometry using Continuous-time Trajectory Optimization

Jan Quenzel and Sven Behnke

Abstract—Simultaneous Localization and Mapping (SLAM) is an essential capability for autonomous robots, but due to high data rates of 3D LiDARs real-time SLAM is challenging. We propose a real-time method for 6D LiDAR odometry. Our approach combines a continuous-time B-Spline trajectory representation with a Gaussian Mixture Model (GMM) formulation to jointly align local multi-resolution surfel maps. Sparse voxel grids and permutohedral lattices ensure fast access to map surfels, and an adaptive resolution selection scheme effectively speeds up registration. A thorough experimental evaluation shows the performance of our approach on multiple datasets and during real-robot experiments.

I. INTRODUCTION

LiDAR plays a major role in environment perception and mapping for autonomous driving [1], unmanned ground vehicles (UGVs) [2] and unmanned aerial vehicles (UAVs) [3]. SLAM and odometry systems provide the basis for many autonomy or assistance functionalities like out-of-sight operation in GNSS-denied environments and reduce strain on the operator while improving their awareness of the surrounding area.

Despite much progress, robustness and reliability in crowded, dynamic scenes and close to structures remain challenging. In recent years, the amount and density of LiDAR measurements increased tremendously which poses new challenges for real-time processing of large point clouds. These factors are essential when a risk-minimizing state, like stopping, is difficult to maintain, e.g., for UAVs.

Most odometry and SLAM systems do not take full advantage of dependencies between consecutive LiDAR scans when aligning a scan against a local map or a previous scan, only jump-starting their registration with prior motion estimates. This may lead to unrealistic jumps in the trajectory since the sensor motion imposes a dependence between consecutive scans. We address this limitation with a continuous-time trajectory representation.

The main contribution of this paper is our novel Multi-Adaptive-Resolution-Surfel (MARS) LiDAR odometry system¹ that jointly registers multiple point clouds against a local multi-resolution surfel map using a continuous-time

This work has been supported by the German Federal Ministry of Education and Research (BMBF) in the project “Kompetenzzentrum: Aufbau des Deutschen Rettungsrobotik-Zentrums (A-DRZ)”

Institute for Computer Science VI, Autonomous Intelligent Systems, University of Bonn, Friedrich-Hirzebruch-Allee 8, 53115 Bonn, Germany, quenzel@ais.uni-bonn.de

¹We will open-source our LiDAR odometry at: https://github.com/AIS-Bonn/lidar_mars_registration. An accompanying video is available at https://ais.uni-bonn.de/videos/iros2021_quenzel/

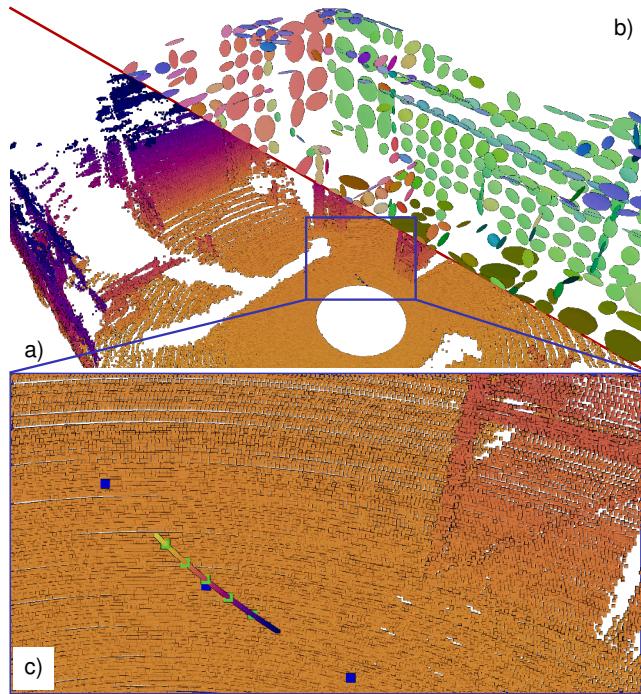


Fig. 1: Multi-resolution surfel maps with adapted resolution represent the point clouds a) during registration with a continuous-time trajectory B-spline. The surfel b) color depends on their normal, while point color depends on height. c) The control points (blue dots) interpolate the scan poses (green dots) and form the spline (blue to yellow).

Lie group B-Spline [4], as visualized in Fig. 1. We adaptively select the most efficient resolution for registration and employ (block-)sparse voxel grids or permutohedral lattices for surfel map storage. We modify the GMM formulation of Droeßel et al. [5] to improve numerical stability and introduce a normal-distance-based weighting. We thoroughly evaluate our system to support our key claims, which are:

- our system provides reliable pose estimates with state-of-the-art quality on a variety of datasets,
- our GMM is more numerically stable and more appropriate given the typical LiDAR sensor geometry,
- our adaptive resolution selection effectively reduces the required computation without degrading accuracy,
- our system runs in real-time onboard a UAV, enabling safe operation in GNSS-denied environments.

II. RELATED WORK

Point cloud registration is a well-researched topic and has wide applicability [6]. A basic registration method is the

Iterative Closest Point (ICP) [7] algorithm. ICP aligns scan and model point clouds in an iterative two-step process. In the first step, the algorithm establishes correspondences between the two point clouds. The second step calculates a transformation to reduce the distance between all corresponding points and repeats both steps with the transformed scan until reaching some optimization criteria. The original formulation assumes perfect correspondences, thus suffers in reality when the sampling locations differ due to a moving sensor.

Another popular scan registration approach is the Normal distribution transform (NDT) [8]. Here, the model point cloud is represented by normal distributions within a regular grid. This reduces the memory consumption as well as computation time for nearest-neighbor searches. NDT aims at maximizing the likelihood of all scan points to observe the underlying surface element (surfel) described by the normal distributions.

Similarly, Segal et al. [9] rephrased ICP within a probabilistic framework that allows incorporating information from the correspondence's covariance. Hence, points on a planar surface will be pulled together in normal direction but have more leeway along the surface.

Once there is a method to register (feature-)points, continuously estimating the sensor pose w.r.t. an updating local map becomes possible. The localization and mapping (LOAM) approach [10] extracts feature points on planar surfaces and edges from the current scan based on the local curvature. Matching features against the previous scan allows estimation of the relative motion at scan frequency. Previous pose estimates help to undistort a newly incoming scan. Every n -th scan is then further processed in the mapping thread and aligned against and integrated into the map. LeGO-LOAM [2] adapted the general approach for horizontally placed LiDARs on UGVs under the assumption of always being able to measure the ground plane. More recently, filter-based approaches [11], [12] use the same feature extraction to directly fuse LiDAR with IMU.

SuMa [13] performs a projection-based data association to avoid the need for costly nearest neighbor associations. For this, it projects the current point cloud from spherical coordinates to an image and renders a model view of the surfel map using OpenGL and the currently estimated pose. This allows easy association between projected points and rendered surfels and enables frame-to-model alignment via ICP with the point-to-plane metric. Afterwards, the map integrates the new scan surfels with an exponentially moving average if they are more accurate. A binary Bayesian Filter estimates the stability and reliability per surfel such that only stable surfels are kept. A pose graph reduces overall drift upon loop-closures and directly deforms the surfel map via the sensor poses.

Instead of a uniform resolution, MRS LaserMaps [5] represents the environment close to the sensor with higher detail. The registration performs Expectation Maximization (EM) of the joint log-likelihood that a scan surfel is an observation of the GMM of the local map. Circular buffers over grid cells

and the fixed number of points stored within each cell enable map shifting to preserve the map's egocentric property. The shifted local map is added to a pose graph to reduce drift over time using constraints between neighboring local maps. The approach was extended in [14] to a hierarchical pose graph and a refinement step to realign scans within a previous local map. These scans were further undistorted by a least-squares fit of a cubic Lie group B-Spline to interpolate between scan poses. In practice, this method provided offline generated maps for robot localization [3].

Elastic LiDAR Fusion [15] uses only a linear continuous-time trajectory. Here, a single transformation linearly interpolates the trajectory within a time segment under an inherent constant-velocity assumption. For the rotation, the Log map of the Lie group $SO(3)$ lifts the relative rotation between start and end pose to its vector-spaced Lie algebra $\mathfrak{so}(3)$, where the interpolation itself takes place. The exponential map Exp maps back to the interpolated rotation. This simple strategy is quite efficient and fast but is limited in practice by the constant-velocity assumption. Their trajectory optimization facilitates geometric constraints penalizing deviations along the normal direction between individual surfels at different time steps within the same scan as well as towards the global map and inertial constraints for rotational velocity and acceleration from IMU. Furthermore, the authors improve map consistency on loop closures through a deformation graph.

Although we base our work upon MRS LaserMaps, our real-time odometry system (Fig. 2) is a full redesign with robustness and efficiency in mind to cope with the large number of scan points generated by modern LiDAR sensors. While most odometry systems align each LiDAR scan individually against the map, we jointly register multiple scans at once in a sliding-window fashion using the continuous-time trajectory B-Spline representation by Sommer et al. [4]. In contrast to MRS LaserMaps, we do not use dense but sparse voxel grids or lattices for each level within our multi-resolution surfel map. Additionally, we scale the surfels' GMM weight to balance the influence of differently sampled areas due to sensor geometry. We adaptively select the appropriate resolution for registration instead of the finest available. Furthermore, we fuse and shift maps via their surfels instead of point-wise and apply a keyframe-based sliding-window for the local map such that we integrate only scans with differing view poses.

III. OUR METHOD

The point cloud P of a scan is a set of points $\mathbf{p}_i \in \mathbb{R}^3$. The statistics of all points falling into a single voxel cell are represented by a surfel with their mean value μ and the covariance Σ . The surfel normal \mathbf{n} is computed as the eigenvector to the smallest eigenvalue of Σ . We regard a surfel as *valid*, if it represents at least 10 points and at least the two largest eigenvalues are non-zero.

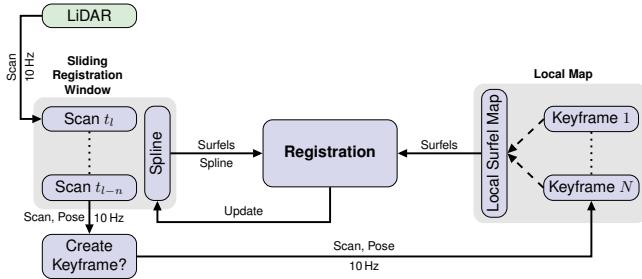


Fig. 2: System overview: A continuous-time trajectory spline describes each scans pose within the current sliding-registration-window. The registration aligns the scan surfels with a local surfel map and updates the spline. Keyframes are added if necessary to the sliding-window of the local map and combined in the local surfel map.

A. Multi-resolution Surfel Map

Our map covers a cubic volume with side length b , e.g., twice the sensor range. The map origin is at the center of this volume.

We subdivide this volume in a uniform volume element (voxel) grid or a 3-dimensional permutohedral lattice with tetrahedral cells. In both cases, we set the distance between adjacent vertices/corners to fixed-size m .

A scaled regular grid $(d+1)\mathbb{Z}^{d+1}$ projected along the one vector $\mathbf{1}$ onto the hyperplane $H_d : \mathbf{p} \cdot \mathbf{1} = 0$ forms the lattice [16]. Each vertex has $2(d+1)$ neighbors and its coordinates sum up to zero. In the lattice, we attribute all points to their closest vertex. A simple rounding algorithm provides the enclosing simplex for each point. One advantage of the lattice compared to the regular voxel grid is the better scalability for higher dimensions. More importantly, a vertex has $2(d+1)$ direct neighboring vertices instead of 3^d , which is more efficient for our soft assignments. Similarly, we do not allocate the whole dense volume but use a sparse lattice or a (block)-sparse voxel grid based on hash-maps. For a more appropriate representation of the sensor geometry, we increase the detail close to the sensor by introducing multiple finer levels. With an increasing level, we half the side length, keep the number of cells constant, and centered at the origin. For registration, the finest available resolution is used or adaptively selected (Sec. III-C).

Within a cell, we have a double-ended queue (deque) to store individual surfels from different scans and a combined surfel. The deque enables easier and faster removal of old scans from the map, in comparison to MRLaserMaps. Adding a new scan creates a new surfel on top of the deque. New scan points are fused only into the new surfel. The cell center is subtracted prior to fusion for numerical stability of the surfel. After processing all points, the changed cells update the combined surfel. During registration, we only use the combined surfels. Fig. 3 shows an example for a single scan as well as the local map with either a voxel grid or the lattice.

Each map stores its pose $T_{ws} \in SE(3)$ and the sensor origin $\mathbf{o} \in \mathbb{R}^3$, since these may change during map fusion. Every scan within the sliding registration window has its

own surfel map, placed at its respective origin. The scan is added at the origin with identity transform. Although this results in differing grid alignment between the scan's surfel map and the local map, it greatly simplifies and speeds up the registration. Grid realignment of all scans within the window is costly. Using the current estimate for the newly added scan would require a more complex spline and would take longer to optimize. Furthermore, we alleviate the grid differences with the soft assignment (Sec. III-B).

We adopt a sliding-window keyframe approach for our local map. Once the distance towards the previous keyframe becomes too large, we add the last scan P from the sliding registration window to the local map. If there are too many keyframes in the local map, we remove the oldest one, including the corresponding surfels from the cells deque and recompute the combined surfel after adding the new scan. The point cloud P in the sensor frame is transformed to the local map frame and then integrated. Before integration, we compute how many cells the new sensor pose moved away from the origin on the coarsest level. Once the difference is above a certain threshold, we shift the whole map to maintain its egocentric property. Shifting by a multiple of the coarsest cell enables efficient swapping of cells instead of cumbersome recomputation from points.

B. Sliding-Window Continuous-time Trajectory Registration

We represent our trajectory using the cumulative B-Spline formulation of Sommer et al. [4] on \mathbb{R}^3 for the translation and on the Lie group $SO(3)$ for the rotation. A B-Spline of order k itself is C^{k-1} continuous and defined by its k control points [4]:

$$i(t) = \lfloor (t - t_0)/\Delta t \rfloor, \quad (1)$$

$$u(t) = ((t - t_0) \bmod \Delta t)/\Delta t, \quad (2)$$

$$\mathbf{p}(u) = \mathbf{p}_i \cdot \prod_{j=1}^{k-1} \lambda_j(u) \cdot (\mathbf{p}_{i+j} - \mathbf{p}_{i+j-1}), \quad (3)$$

$$R(u) = R_i \cdot \prod_{j=1}^{k-1} \text{Exp}(\lambda_j(u) \cdot \mathbf{d}_j^i), \quad (4)$$

$$\mathbf{d}_j^i = \text{Log}(R_{i+j-1}^{-1} R_{i+j}) \in \mathbb{R}^3, \quad (5)$$

$$T_X(t) = \begin{bmatrix} R(u(t)) & \mathbf{p}(u(t)) \\ \mathbf{0} & 1 \end{bmatrix}. \quad (6)$$

Here, i is the index of the first control point in the active time segment $[t_i, t_{i+1})$ and u is the normalized time since the start of that segment. For $R \in SO(3)$ Log maps from the Lie group to its vector-spaced Lie algebra and Exp maps back to $SO(3)$. Although the control points $X_i = (R_i, p_i) \in SO(3) \times \mathbb{R}^3$ are from two different Splines, the result $T(t)$ is a rigid transform in $SE(3)$.

In practice, we set the order to $k = 3$ and thus, have to optimize all k control points during registration — independent of the number of scans n . The time interval Δt is adaptively chosen such that the last scan l in the sliding-window is within the current time interval $[t_{l-n}, t_l + t_{pred}]$. Here, t_{l-n} is the time of the last shifted out scan. We update the control

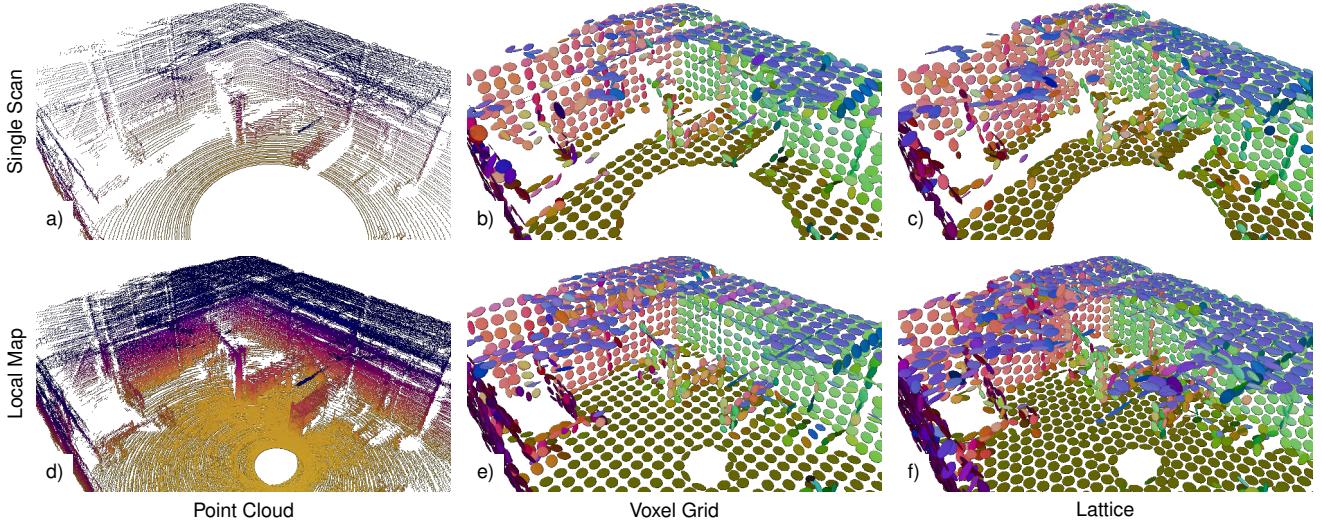


Fig. 3: Comparison between surfel maps with voxel grid [b,e] and permutohedral lattice [c,f].

points after shifting through the following optimization:

$$X_{init} = \arg \min_X \sum_{i=1}^n \mathbf{e}_p^\top \mathbf{e}_p + \mathbf{e}_R^\top \mathbf{e}_R, \quad (7)$$

$$\mathbf{e}_p = \mathbf{p}(u(t_j)) - \mathbf{p}_{j-1}, \quad (8)$$

$$\mathbf{e}_R = \text{Log}(R_{j-1}^{-1} R(u(t_j))), \quad (9)$$

$$j = l - n + i, \quad (10)$$

where R_{j-1} and \mathbf{p}_{j-1} are the previous estimates for the unshifted spline. The prediction time offset t_{pred} helps during the deployment of the system on real robots to predict the robot's position, e.g., 0.1 s or one scan into the future, effectively reducing the lag due to processing time.

We model the likelihood of a scene surfel s observing a map surfel m as the following normal distribution:

$$\mathbf{e}_{sm}(T) \sim \mathcal{N}(\mathbf{d}_{sm}, \Sigma_{sm} + \sigma^2 I), \quad (11)$$

$$\mathbf{d}_{sm}(T) = T\mu_s - \mu_m, \quad (12)$$

$$\Sigma_{sm}(T) = \Sigma_m + R\Sigma_s R^\top, \quad (13)$$

with R being the rotation of the rigid transform T and a resolution-depending scaling term σ^2 . For better readability, we will drop the transformation argument for above formulas whenever possible. This model would require a hard decision whether or not s corresponds to m . Instead, we apply a GMM that represents a mixture of multiple map surfels being observed by s following the normal distribution (11) with an additional uniform component $p(o_s)$ for outliers:

$$p_s(T) = p(o_s) + \sum_{m \in \mathcal{A}_s} p(a_{sm}) p(\delta_{sm}) p(\mathbf{e}_{sm}), \quad (14)$$

$$p(o_s) = p(o) p(\mathcal{N}(\mathbf{0}, R\Sigma_s R^\top + \sigma^2 I)). \quad (15)$$

Here, \mathcal{A}_s is the set of map associations for surfel s with the prior association likelihood $p(a_{sm})$. For more details on the GMM, we refer the reader to [5].

$p(\delta_{sm})$ further takes the similarity between associated surfels w.r.t. normal and viewing direction into account.

Both are modeled as:

$$p_v(\mathbf{v}) \sim \mathcal{N}(\arccos((R\mathbf{v}_s) \cdot \mathbf{v}_m), (\pi/8)^2). \quad (16)$$

Additionally, we model the distance in the normal direction:

$$d_n \sim \mathcal{N}(\mathbf{n}_m^\top \Sigma_{sm}^{-1} \mathbf{d}_{sm}, \sigma^2), \quad (17)$$

assuming independence results in the approximation:

$$p(\delta_{sm}) = p_v(\mathbf{n}) p_v(\mathbf{v}) p(d_n). \quad (18)$$

Given this GMM formulation, we seek to find the spline control points X that maximize the logarithm of the joint observation likelihood over the mixture:

$$X^* = \arg \max_X \sum_{i=1}^n \sum_{s \in \mathcal{S}} \log(p_s(T_X(t_{l-n+i}))), \quad (19)$$

and solve (19) with EM. In the E-Step, we establish associations for all surfels within the sliding-window. The pose T_c , evaluated the spline at time t_c , transforms the mean μ_s from the scan's sensor frame into the map frame. Then, we lookup the corresponding surfel in the map and check its 1-hop-neighborhood for valid surfels. In 3D, this creates up to 27 scan-map associations for a 3D voxel grid or 9 for the lattice. We calculate per association their conditional likelihood w_{sm} given the current estimate T_c :

$$w_{sm}(T_c) = \frac{p(a_{sm}) p(d_{sm}) p(\mathbf{e}_{sm})}{p_s(T_c)}, \quad (20)$$

and keep these associations and weights fixed during the M-Step to compute an updated estimate for the control points X :

$$X^* = \arg \min_X \sum_{i=1}^n \sum_{s \in \mathcal{S}_i} \sum_{m \in \mathcal{A}_s} r_{sm}(T_X(t_{l-n+i})), \quad (21)$$

$$r_{sm}(T_c) = w \mathbf{d}^\top \Sigma_{sm}^{-1} \mathbf{d}. \quad (22)$$

The Levenberg-Marquardt algorithm optimizes (21).

We found empirically that the GMM assigns a higher weight w_{sm} to surfels in the 3D-LiDAR's vicinity. This

leads in situations where one translation direction was underconstrained, e.g. open park areas with tree trunks further away, or dynamic environments with moving objects close to the sensor, to wrongly estimate the translation as staying in place while the orientation was correct. Thus, following HeRO [17], we analyzed the condition number κ of the covariance of surfel normals — since κ reflects the difficulty to accurately solve a linear system of equations. Weighting the normals with their conditional likelihood w_{sm} increases κ — in some cases by a factor of up to 50. We found that this relates to the higher measurement density for surfaces close to the LiDAR, because the prior association likelihood $p(a_{sm})$ directly incorporates the number of measurements per surfel and the LiDAR samples the environment non-uniformly in 3D. Hence, we inversely weight w_{sm} by the number of surfel measurements to effectively level the influence between far and close surfels.

C. Adaptive Resolution Selection

Always restricting to the finest available map resolution for registration is computationally inefficient for planar surfaces like walls, roads, floors, or ceilings. Hence, we select the most appropriate resolution adaptively.

We start collecting valid surfels on the finest map scale and check for not yet processed valid surfels for the following three conditions on the normalized Eigenvalues $\lambda_0 \leq \lambda_1 \leq \lambda_2 \in \mathbb{R}$ s.t. $\sum_i \lambda_i = 1$ of the surfel covariance matrix C :

$$\lambda_0 < \theta_{planar}, \quad (23)$$

$$\lambda_0 < \lambda_1 \theta_{scale}, \quad (24)$$

$$\lambda_1 < \theta_{degenerate}. \quad (25)$$

The first two cases directly relate to planar surfaces, while the last case often occurs for surfels created from a single scan line. Fig. 4 illustrates these. If at least one condition is true, we test all valid neighbors falling into the same coarser surfel and calculate the mean normal vector $\bar{\mathbf{n}}$. Once all valid finer surfels pass this test, we perform the same check on the coarser surfel. Additionally, we ensure that the mean and coarser normal are similar:

$$|\mathbf{n}_c \cdot \bar{\mathbf{n}}| > \theta_n. \quad (26)$$

Upon passing these tests, the coarser surfel will be further processed instead of the smaller ones, thus, reducing the total number of surfels and speeding up registration. Typical values for the thresholds are $\theta_{planar} = 0.01$, $\theta_{scale} = 0.01$, $\theta_{degenerate} = 0.1$, $\theta_n = 0.8$.

IV. EVALUATION

We evaluate our method on the Urban Loco dataset [18] for autonomous driving, the Newer College dataset [19], and nine self-recorded UAV flights through the DRZ Living Lab. Furthermore, we show qualitative results from multiple UAV flights. All experiments were conducted on a laptop with an Intel Core i7-6700HQ with 32 GB of RAM and an NVIDIA GeForce GTX 960M. We compare our method

against the two LOAM variants A-LOAM² and F-LOAM³ as well as SuMa⁴. We modified A/F-LOAM to use ROS *message_filters*⁵ to enforce real-time with a limited queue size of 10 to prevent processing too old information. Furthermore, we deactivated for fair comparison the loop-closing of SuMa. We did not include MRSLaserMaps since it lost track on all tested sequences. Our method is implemented in C++ and runs on the CPU only. In all experiments we perform up to three iterations within the Levenberg-Marquardt algorithm to optimize (21). The storage type is by default set to a sparse voxel grid unless otherwise noted.

For all systems, we evaluate the root mean squared absolute trajectory error (ATE) for the at-runtime estimated poses against the provided ground-truth. The best result will be in bold, while the second best will be colored blue. For both LOAM-derivatives, we report the more accurate optimized poses instead of the initial odometry poses and ignore unoptimized scans during ATE computation.

A. Newer College Dataset

Ramezani et al. [19] replicated the route of the New College Dataset with a handheld Ouster OS-1 LiDAR with 64 beams at 10Hz. We selected the sequences “01_short” and “02_long” and “05_quad_with_dynamics”. The first two experiment runs “01” and “02” consist of multiple loops between buildings and through a park for over 1530s and 2656s. Sequence “05_quad_with_dynamics” is a shorter 398s run with faster and partially swinging motion during four loops within a quad. The authors aligned the LiDAR scans against a terrestrial laser scanner to provide ground-truth poses. Tab. I shows the resulting ATE. Although the LOAM-derivates’ odometry processed each scan, both could only fully optimize every third scan in real-time. On the second sequence, SuMa unfortunately lost track in the park area while A-LOAM and F-LOAM exhibited drift during orientation changes. For sequence “05” the limited size of the quad and A-LOAMs’ local map optimization allows A-LOAM to relocalize, thus keeping the error low. In contrast, our sliding-keyframe-window accumulates more drift over time but maintains high accuracy such that in the future the extraction of loop-closure candidates based on some vicinity criterion is possible.

B. Urban Loco Dataset

Wen et al. [18] equipped two cars with LiDAR and other sensors to capture highly urbanized areas throughout San Francisco and Hong Kong. A navigation system with RTK-GPS and IMU provides the ground-truth for all sequences. The Coli Tower sequence (ULCT) took 248s and is a 1.8 km long drive up hill within a dynamic environment. Similarly, the Lombard Street sequence (ULLS) is a 1 km drive and took 253s. Both were captured with a RS-LiDAR-32 while

²<https://github.com/HKUST-Aerial-Robotics/A-LOAM>

³<https://github.com/wh200720041/floam>

⁴<https://github.com/jbehley/SuMa>

⁵http://wiki.ros.org/message_filters

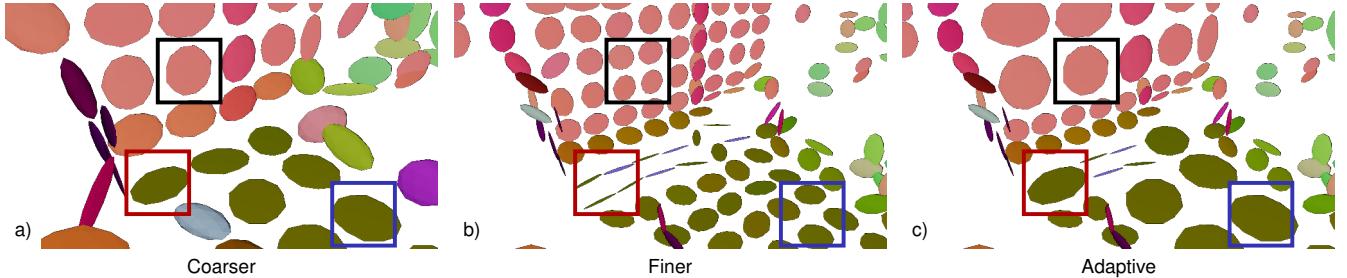


Fig. 4: Adaptive resolution selection selects coarser surfel in areas of finer resolution with planar surfels. The red rectangle highlights merging multiple degenerate surfels, while the black and blue are examples for Eq. 23 and Eq. 24.

TABLE I: ATE [m] Evaluation on scenes from the Newer College (NC) and Urban Loco (UL) datasets

	Ours (Grid)	Ours (Lattice)	A-LOAM	F-LOAM	SuMa
NC01	2.17229	1.97836	3.30768	101.899	2.04814
NC02	4.93553	5.11236	62.64240	87.08770	X
NC05	0.46826	0.41505	0.14824	2.81843	1.87838
ULCT	4.99616	4.96610	10.37230	6.05617	7.77146
ULLS	7.82386	7.57318	8.40002	7.56019	10.69550
ULHH	2.35564	2.40788	2.43283	2.42415	2.32560
ULLL	2.73179	2.23304	2.07094	2.31379	8.44114
ULSL	3.19269	3.26606	3.16574	3.27051	2.87772
ULWH	2.95064	2.99434	2.42352	2.25923	3.50044
ULT2	2.54422	1.44114	17.34920	2.42821	2.21693
ULT3	1.59692	1.55525	9.67000	2.02969	2.94482
ULH5	1.55182	1.50516	17.35210	2.42821	2.21693

the seven Hong Kong sequences used a Velodyne HDL-32E LiDAR. The Hong Kong sequences took between 150 s and 365 s and are up to 2 km long through dense urban environment.

Tab. I shows the results for all four algorithms. Our method achieves consistent state-of-the-art results with placing first in three and second in another two out of nine sequences.

C. DRZ Living Lab

We collected LiDAR scans from an Ouster OS-0 with 128 beams at 10 Hz attached to a DJI M210 v2 while flying through the DRZ Living Lab⁶. A Motion Capture (MoCap) system provides ground-truth poses in the starting section of the lab, where we recorded multiple flights with up to 2.5 min flight time. All sequences except for the second dataset remain within the MoCap volume with varying angular velocity and linear acceleration. In the second dataset we traversed back and forth through the Living Lab. The flight starts and finishes within the MoCap. Three further experiments were conducted with three persons moving around with a slow (3P-S), medium (3P-M) and fast (3P-F) flying UAV. For comparison, we conducted similar exploration flights with different motion speeds in a static environment. For fair evaluation, we use the UAVs' IMU orientation to compensate for the rotation distortion and supply these scans to all tested methods. Tab. II details the number of scans, the

median norm of the linear acceleration as well as the median and maximum norm of the angular velocity and shows the results for all algorithms.

The surfel-based methods cope well with the faster moving sequences, while the LOAM-derivatives struggle, even with the undistorted scans. Our method performs very well in comparison, with the exception of the “hall”-sequence. This has likely to do with the local map size of all methods, since our sliding-keyframe-window is the only one that did not cover the whole Living Lab.

To analyze the impact of different spline settings, we evaluate multiple combinations on the most challenging sequence “F2”. Tab. III reports the parameters for number of control points k , the number of jointly optimized scans n and the resulting ATE with the actual computation time during registration. The storage type was set to sparse grid since timings for block-sparse and sparse were similar due to the relatively low surfel count. Increasing n provided no benefit, but slows down computation and required in some cases more iterations to compute a stable result. In comparison, the registration took 29.82 ms when using a permutohedral grid with $k = 3$ and $n = 3$ without adaptive resolution selection at an ATE of 0.09489 m. The resolution selection reduced runtime to an avg. 22.95 ms, although the ATE increased slightly to 0.10016 m. In contrast, the voxel grid without adaptive selection could not correctly estimate all rotations resulting in a 90° drift during a high angular velocity maneuver. Similarly, we typically obtained more consistent and accurate results with enabled adaptive resolution selection in open environments.

Fig. 5 shows the aggregated point cloud from the DRZ Living Lab traverse flight.

D. Qualitative UAV Experiment

An early development version of our registration provided the onboard LiDAR odometry for multiple autonomous flights of the same DJI M210v2 UAV through GNSS-denied areas. An EKF fused our pose estimates with the aircraft's IMU. For more details on the autonomous UAV flights, we refer the reader to [20]⁷. Fig. 6 shows the finest resolution of the local surfel map and the aggregated point cloud.

⁶<https://rettungsrobotik.de/living-lab/>

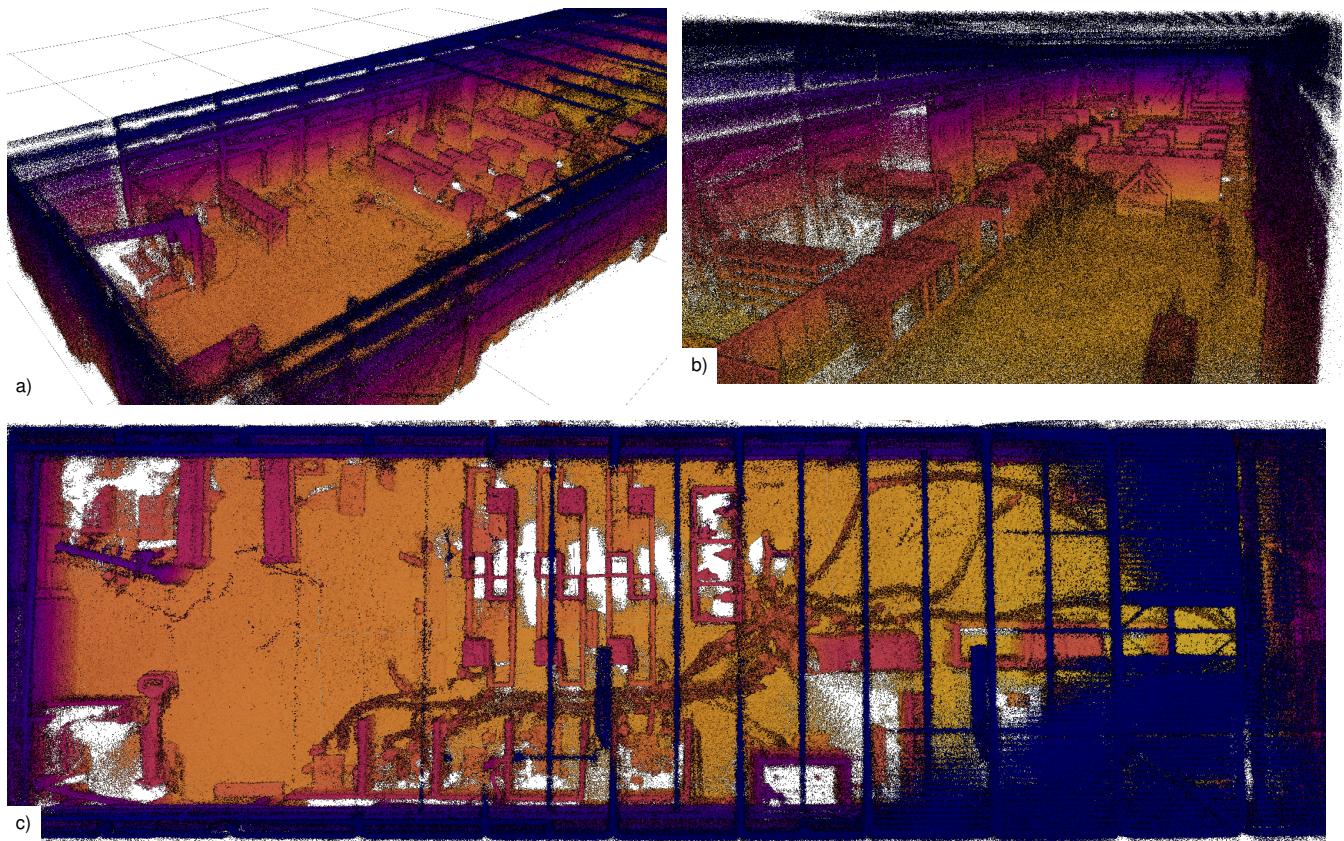


Fig. 5: Aggregated point cloud from the traverse through the DRZ Living Lab. The roof is partially removed for better visualization.

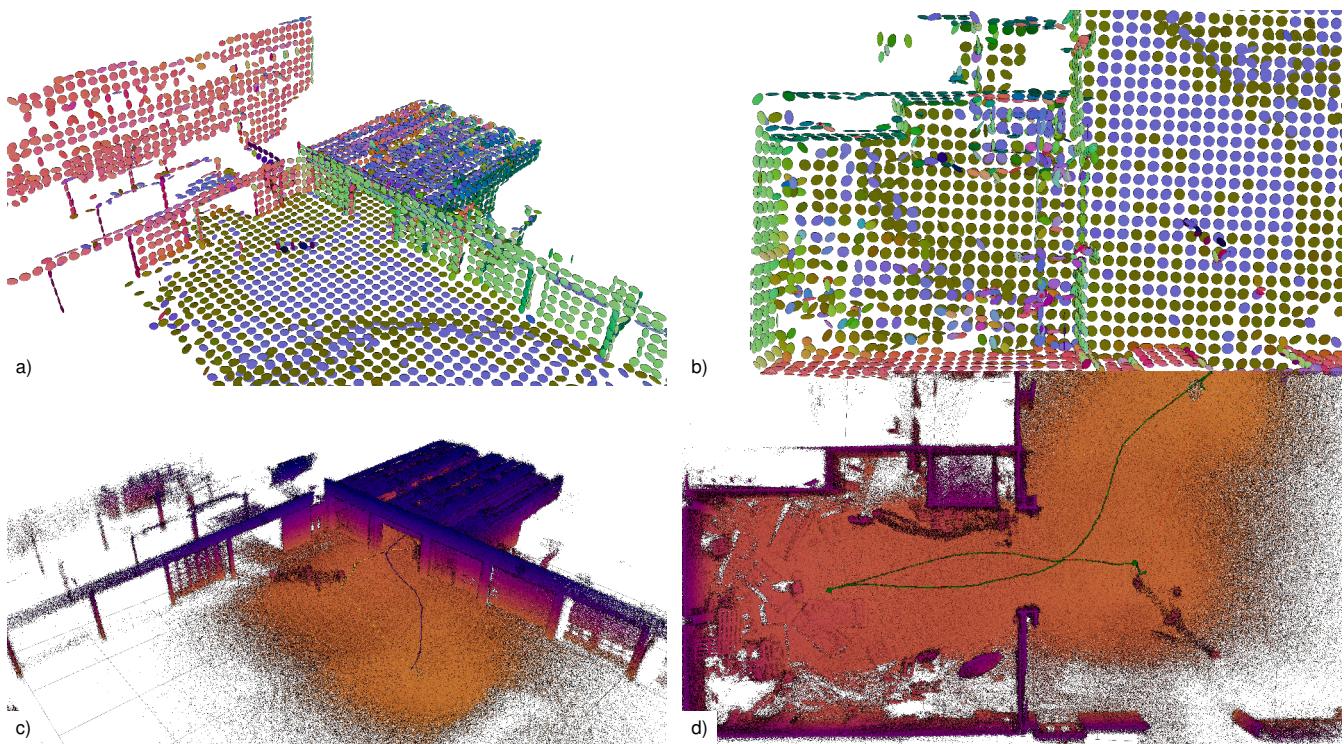


Fig. 6: Local map [a),b)] and aggregated point cloud [c),d)] and from the autonomous flight. The trajectory is colored green in c) and d).

TABLE II: ATE [m] Evaluation for the DRZ Living Lab dataset

Seq.	Scans	median ($\ \alpha\ $) [m/s ²]	median ($\ \omega\ $) [rad/s]	max($\ \omega\ $) [rad/s]	Ours (Grid)	Ours (Lattice)	A-LOAM	F-LOAM	SuMa
Fast	1479	1.13119	0.22876	1.89359	0.094848	0.05566	2.26457	2.11024	0.06370
Hall	2421	0.39893	0.09758	1.14679	0.20538	0.09212	0.03032	0.03850	0.07256
3P-S	1085	0.27435	0.07900	0.86189	0.02181	0.01465	0.02244	0.01665	0.03834
3P-M	650	0.58921	0.11351	1.08112	0.02873	0.02393	0.02517	0.03076	0.04032
3P-F	604	1.99022	0.52543	3.54729	0.06066	0.04995	0.71797	0.83169	0.09472
S1	826	0.48876	0.10443	1.51561	0.04398	0.03968	0.04555	0.04636	0.041957
M1	1458	0.90011	0.21923	1.97838	0.08220	0.06385	3.73529	3.62154	0.17566
F2	795	2.30749	0.48656	3.98107	0.10458	0.10016	2.62554	2.80373	0.08839
F3	957	1.04914	0.24944	1.63187	0.08520	0.07374	0.14406	2.45972	0.17991

TABLE III: Statistics for varying spline parameters on the “F2” sequence in the DRZ Living Lab. Entries with * required five instead of three iterations.

k	Spline		ATE [m]	avg. time [ms]
	n			
2	2		0.10628	41.28
	3		0.10970*	65.45
	4		0.22468*	71.03
3	3		0.10458	50.57
	4		0.10759*	82.53
	5		0.11229*	91.85
4	4		0.11133	117.03
	5		0.11246	121.88
	6		0.12333	121.88

V. CONCLUSION

We presented MARS, a novel multi-resolution surfel-based LiDAR odometry. A sparse permutohedral lattice or voxel grid stores the surfels within our multi-resolution surfel map. Multiple LiDAR scans are jointly registered against the local map using a continuous-time trajectory with adaptively selected surfel resolution. In the future, we plan to incorporate IMU measurements and visual features in the registration process and extend our system with a pose graph to further reduce the drift over time, incorporating GPS poses and loop closures.

REFERENCES

- [1] A. Geiger, P. Lenz, and R. Urtasun, “Are we ready for autonomous driving? the KITTI vision benchmark suite,” in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2012.
- [2] T. Shan and B. Englöt, “LeGO-LOAM: Lightweight and ground-optimized lidar odometry and mapping on variable terrain,” in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2018.
- [3] M. Beul, D. Droeschel, M. Nieuwenhuisen, J. Quenzel, S. Houben, and S. Behnke, “Fast autonomous flight in warehouses for inventory applications,” in *IEEE Robotics and Automation Letters (RA-L)*, 2018.
- [4] C. Sommer, V. Usenko, D. Schubert, N. Demmel, and D. Cremers, “Efficient derivative computation for cumulative B-Splines on Lie groups,” in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020.
- [5] D. Droeschel, M. Schwarz, and S. Behnke, “Continuous mapping and localization for autonomous navigation in rough terrain using a 3D laser scanner,” *Journal of Robotics and Autonomous Systems*, vol. 88, pp. 104–115, 2017.
- [6] D. Holz, A. E. Ichim, F. Tombari, R. B. Rusu, and S. Behnke, “Registration with the point cloud library: A modular framework for aligning in 3-D,” *IEEE Robotics & Automation Magazine*, vol. 22, no. 4, pp. 110–124, 2015.
- [7] P. Besl and N. McKay, “Method for registration of 3D shapes,” in *Sensor Fusion IV: Control Paradigms and Data Structures*, 1992.
- [8] M. Magnusson, A. Lilienthal, and T. Duckett, “Scan registration for autonomous mining vehicles using 3D-NDT,” *Journal of Field Robotics (JFR)*, vol. 24, no. 10, pp. 803–827, 2007.
- [9] A. Segal, D. Haehnel, and S. Thrun, “Generalized-ICP,” in *Robotics: Science and Systems (RSS)*, 2009.
- [10] J. Zhang and S. Singh, “LOAM: Lidar odometry and mapping in real-time,” in *Robotics: Science and Systems (RSS)*, 2014.
- [11] C. Qin, H. Ye, C. Pranata, J. Han, S. Zhang, and M. Liu, “LINS: A lidar-inertial state estimator for robust and efficient navigation,” in *IEEE International Conference on Robotics and Automation (ICRA)*, 2020.
- [12] H. Ye, Y. Chen, and M. Liu, “Tightly coupled 3D lidar inertial odometry and mapping,” in *IEEE International Conference on Robotics and Automation (ICRA)*, 2019.
- [13] J. Behley and C. Stachniss, “Efficient surfel-based SLAM using 3D laser range data in urban environments,” in *Robotics: Science and Systems (RSS)*, 2018.
- [14] D. Droeschel and S. Behnke, “Efficient continuous-time slam for 3D lidar-based online mapping,” in *IEEE International Conference on Robotics and Automation (ICRA)*, 2018.
- [15] C. Park, P. Moghadam, S. Kim, A. Elfes, C. Fookes, and S. Sridharan, “Elastic LiDAR fusion: Dense map-centric continuous-time SLAM,” in *IEEE International Conference on Robotics and Automation (ICRA)*, 2018.
- [16] R. A. Rosu, P. Schütt, J. Quenzel, and S. Behnke, “LatticeNet: Fast point cloud segmentation using permutohedral lattices,” in *Robotics: Science and Systems (RSS)*, 2020.
- [17] A. Santamaría-Navarro, R. Thakker, D. D. Fan, B. Morrell, and A. Agha-mohammadi, “Towards resilient autonomous navigation of drones,” 2020. [Online]. Available: <https://arxiv.org/abs/2008.09679>
- [18] W. Wen, Y. Zhou, G. Zhang, S. Fahandezh-Saadi, X. Bai, W. Zhan, M. Tomizuka, and L. T. Hsu, “UrbanLoco: A full sensor suite dataset for mapping and localization in urban scenes,” in *IEEE International Conference on Robotics and Automation (ICRA)*, 2020.
- [19] M. Ramezani, Y. Wang, M. Camurri, D. Wisth, M. Mattamala, and M. Fallon, “The Newer College dataset: Handheld LiDAR, inertial and vision with ground truth,” in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2020.
- [20] D. Schleich, M. Beul, J. Quenzel, and S. Behnke, “Autonomous flight in unknown GNSS-denied environments for disaster examination,” in *International Conference on Unmanned Aircraft Systems (ICUAS)*, 2021.