

Fast Global Point Cloud Registration using Semantic NDT

Robert Schirmer

Narunas Vaskevicius

Peter Biber

Cyrill Stachniss

Abstract—Robust and accurate point cloud registration is an essential part of many robotic tasks such as SLAM or object pose retrieval. In this paper, we address the problem of global 3D point cloud registration, i.e., the task of estimating the 3D rigid body transform between a source and a target point cloud without any initial guess. Typically, the problem is solved by extracting and matching features to find a data association and then computing a transform that minimizes the squared distance between points. Our approach combines the normal distributions transform and oriented point pair framework and introduces the NDT distance histogram to quickly generate and test candidate transforms. Our method further exploits semantic information if available for greater speed. We implement our algorithm in C++ and compare it to other state-of-the-art approaches on a diverse set of environments. Our evaluation shows that our method outperforms the other approaches, especially concerning run-time and compute efficiency.

I. INTRODUCTION

Point cloud registration is an essential building block for robot applications and key for sensor odometry estimation, SLAM and 3D object reconstruction. In graph-based SLAM, the registration problem is often formulated in a local and a global setting. Local registration is used in LiDAR odometry to accurately track the robot pose and relies on an initial transform estimate. Global registration handles any relative motion and is used in loop closure estimation to correct the accumulated drift and generate a globally consistent map.

In this paper, we consider the problem of global 3D point cloud registration and present an approach that achieves fast and at the same time precise results. The challenges of the problem arise from spatial aliasing, partial overlap between both point clouds, the noise and outliers due to the dynamics of the scene or large differences from one viewpoint to another due to sensor geometry.

Approaches for global 3D point cloud registration often build on descriptor extraction and matching processes. In practice, computing hand-crafted features such as fast point feature histograms [27] for large inputs may be expensive, parameter sensitive or yield unsatisfactory performance. In contrast, deep learnt features such as fully convolutional geometric features [9] perform robustly and have faster processing speeds, but require training data and a GPU.

Robert Schirmer, Narunas Vaskevicius and Peter Biber are with Robert Bosch GmbH. Robert Schirmer and Cyrill Stachniss are with the Center for Robotics, University of Bonn, Germany. Cyrill Stachniss is additionally with the Lamarr Institute for Machine Learning and Artificial Intelligence, Germany.

This work has partially been funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy, EXC-2070 – 390732324 – PhenoRob and under STA 1051/5-1 within the FOR 5351 (AID4Crops) and by the European Union's Horizon Europe research and innovation program under grant agreement No 101070405 (DigiForest).

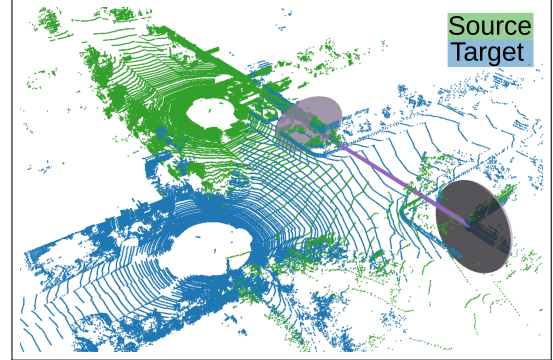


Fig. 1: Our approach searches for a corresponding pair of NDTs (purple) in source (green) and target (blue) to solve the global point cloud registration problem.

The main contribution of this paper is a novel approach for global registration which is fast, robust, and leverages semantic information as given by semantic segmentation. Our algorithm builds on three approaches. We use the normal distributions transform (NDT) proposed by Biber et al. [5] for voxelizing the point clouds, computing normals, and estimating a registration score. We generate candidate transforms by sampling point pairs and their normals using a method inspired by Winkelbach et al. [34] which we guide by introducing the NDT distance histogram. We accelerate the fitness assessment of each candidate transform using the bail-out test by Capel [6]. Our experimental evaluation shows that even without semantic information, our global matcher performs on par with the state of the art and is able to match difficult registration problems such as the one shown in Fig. 1. We make three key claims in this paper: Our global 3D point cloud registration approach is able to (i) perform strongly across different settings; (ii) generate results faster than the state of the art; (iii) optionally leverage semantic information for faster results. These claims are backed up by the paper and our experimental evaluation.

II. RELATED WORK

We point towards the survey by Huang et al. [14] for a general discussion on point cloud registration as well as the one by Yin et al. [37] on LiDAR-based place recognition. Here, we mainly focus on correspondence-based methods, dense methods, and the exploitation of semantic information for point cloud registration.

Correspondence-based methods have descriptor extraction and matching steps. A prominent hand-crafted descriptor is fast point feature histograms (FPFH) by Rusu et al. [27], where local geometry is encoded as a histogram of neigh-

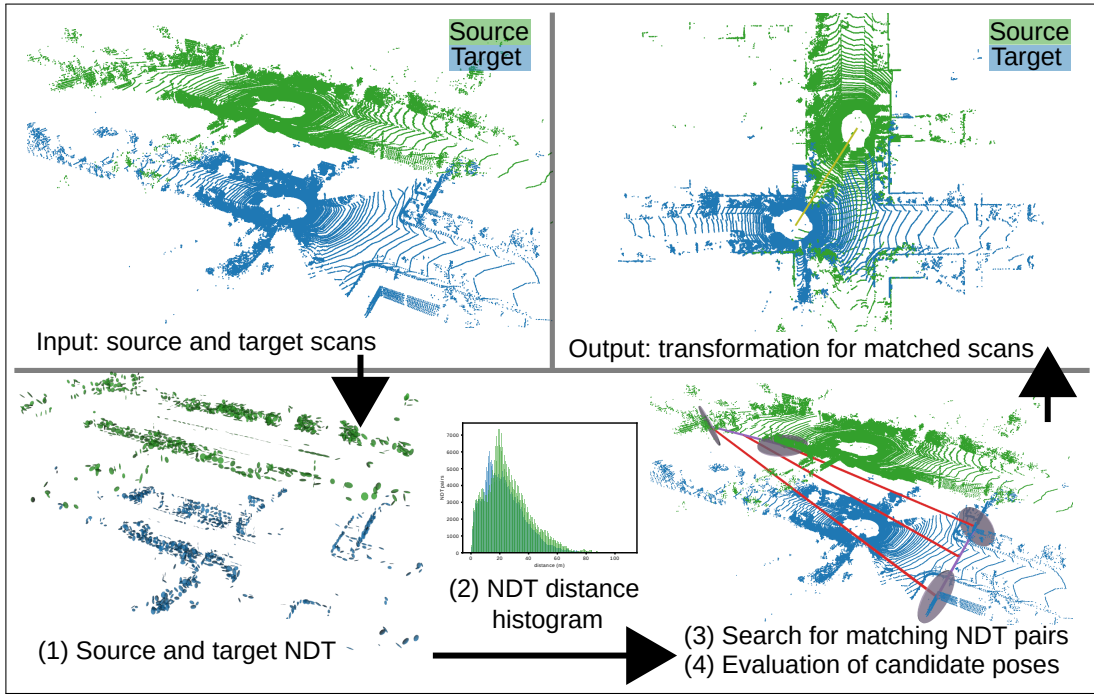


Fig. 2: Overview of our approach. To match source (green) with target (blue) we (1) compute the NDTs, (2) generate the NDT distance histogram, which enumerates and references by distance all possible NDT pairs, (3) search for corresponding NDT pairs (purple) by using the NDT distance histogram and (4) evaluate the hypothesized transforms (red) established by aligning the NDT means and normals.

boring points and normals. A lot of recent work has focused on data driven descriptor learning such as 3DMatch by Zheng et al. [40] or FCGF by Choy et al. [9]. Learnt descriptors can outperform hand-crafted ones, but require a GPU and may drop in performance when transferring between domains as discussed by Drory et al. [10]. Some methods exploit a bird’s eye view representation and density maps to align point clouds, also for finding loop closures [13]. One may also exploit sequence of sensor readings [33].

After obtaining the descriptors, correspondences between them are established and matched to extract a relative motion between both point clouds. A popular family of approaches is based on RANSAC, which works by repeatedly sampling a set of point matches, estimating a motion, and calculating the score as the fraction of point matches agreeing with the motion. RANSAC has been extended with improvements to sample selection as done by Barath et al. [3], or early rejection of non-promising candidates such as Matas et al. [20] and Capel [6]. As randomized approaches converge slowly in the presence of high outlier rates, recent methods have proposed more robust and deterministic descriptor matching. TEASER++ by Yang et al. [36] formulates the problem as a graph and uses robust maximum clique methods to match the descriptors. Zhang et al. [41] extend this to use maximal cliques and combine their approach with deep-learnt methods to achieve state-of-the-art results.

Dense methods without descriptor extraction have also been proposed. In principle, a rigid transformation can be estimated from three point correspondences between the target and source point clouds. The correspondence search can be simplified to four point congruent sets as proposed

by Aiger et al. [1] and Mellado et al. [22]. Winkelbach et al. [34] present an approach for global registration based on oriented point (position and normal) pairs. This is further extended by Papazov et al. [24] to 3D object identification and pose estimation.

Lim et al. [16], [17] discuss the degeneracy problem which occurs when outlier rejection prunes too many correspondences. Their Quatro extension of TEASER++ uses the Atlanta assumption to only estimate the yaw rotation angle during point cloud registration, as in many applications roll and pitch are known from an IMU. We note that by construction, dense methods are unaffected by degeneracy as the problem geometry is not abstracted into a descriptor-matching problem.

Zaganidis et al. [39] use semantic segmentation in the data association step of NDT to achieve good results in global registration settings. Semantic segmentation is also used by Chen et al. [7] in SUMA++ to achieve highly accurate results in the KITTI odometry benchmark [12]. Yin et al. [38] extend TEASER++ [35] with semantic information and present strong registration results, also with noisy semantic labels.

The great performance of oriented point pair and semantically assisted methods have inspired us to build upon them and introduce several key improvements, which taken together are the novelty we present in this paper. First, we integrate the oriented point pair approach into the NDT framework. This changes the main transform estimation primitive from corresponding point pairs with normals to corresponding NDT pairs. This also enables us to efficiently exploit local shape information by using the NDT-D2D distance formulation when evaluating a transform. The second

key improvement we present is the introduction of the NDT distance histogram to optimize candidate pose extraction. This preprocessing step guides the search for corresponding NDT pairs towards the most promising ones. Finally, our method's structure enables the effective use of pixel-wise semantic information: we use it to reduce the quadratic cost of computing the NDT distance histogram, and to further semantically focus the search for corresponding NDT pairs.

III. OUR APPROACH

Let P be the source and Q the target point cloud. The goal of global registration is to find a rigid 3D isometry transformation $T_P^Q \in \mathbf{SE}(3)$ that aligns P to Q s.t. the squared distance between corresponding points is minimized.

$$T_P^Q = \underset{T}{\operatorname{argmin}} \sum_{(\mathbf{p}, \mathbf{q}) \in K(\tau_t)} \|T\mathbf{p} - \mathbf{q}\|_2, \quad (1)$$

with $\mathbf{p} \in P$, $\mathbf{q} \in Q$ and K being the set of nearest neighbor correspondences with a distance smaller than τ_t . Our method uses four steps to find T_P^Q , as illustrated in Fig. 2: (1) NDT computation of the source and target point clouds, (2) generation of the NDT distance histogram, (3) search for corresponding NDT pairs in source and target, and (4) evaluation of hypothesized transforms established by aligning the NDT means and normals.

A. NDT Computation

In the first step, we compute the normal distributions transform of both P and Q , $\text{NDT}(P)$ and $\text{NDT}(Q)$. NDTs are useful for point cloud representation and registration and have been introduced by Biber et al. [5] for 2D registration and extended to 3D by Magnusson et al. [19]. We define $\text{NDT}(P)$ as the set of all NDT_i obtained from segmenting P into axis-aligned voxel cells V of fixed size, and fitting a 3D Gaussian distribution to the points that fall within each voxel. Let S_i be the set of points in voxel V_i , then NDT_i with mean μ_i and covariance C_i are defined as:

$$\text{NDT}(P) := \{(\text{NDT}_0, \dots, \text{NDT}_i) \mid \forall i \in V\} \quad (2)$$

$$\text{NDT}_i := (\mu_i, C_i) \quad (3)$$

$$\mu_i = \frac{1}{|S_i|} \sum_{\mathbf{p} \in S_i} \mathbf{p} \quad (4)$$

$$C_i = \frac{1}{|S_i| - 1} \sum_{\mathbf{p} \in S_i} (\mathbf{p} - \mu_i)(\mathbf{p} - \mu_i)^\top \quad (5)$$

We further compute the Eigenvectors and corresponding Eigenvalues by computing the Eigen decomposition of the covariance matrix C_i , and avoid nearly singular matrices following Magnusson [18].

B. Candidate Transform Extraction

In the second step, we generate candidate transforms by adapting the oriented point pair approach by Winkelbach et al. [34] to NDTs and extending it with the NDT distance histogram. For this, we describe the relationship between

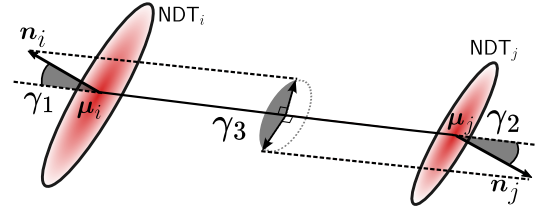


Fig. 3: Angles used to describe the relationship between NDT pairs.

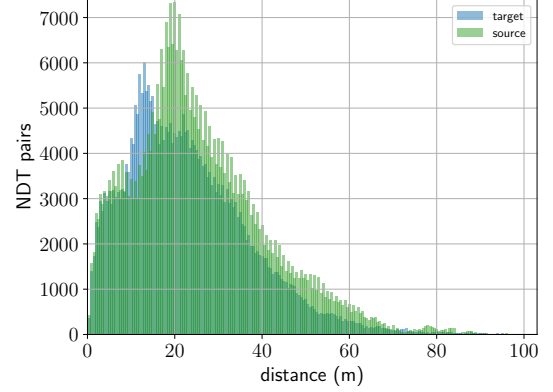


Fig. 4: The NDT of a typical KITTI scan yields approximately 10^3 cells and 10^6 cell pairs. The NDT distance histogram enumerates and indexes all pairs by their distance.

NDT cell pairs with their means μ and normals \mathbf{n} . The normal \mathbf{n} is the Eigenvector associated with the smallest Eigenvalue of the covariance C , multiplied when needed by -1 s.t. it faces outwards from the center of the pair. We denote two NDT pairs, $(\text{NDT}_i, \text{NDT}_j) \in \text{NDT}(P)$ and $(\text{NDT}_k, \text{NDT}_l) \in \text{NDT}(Q)$, to correspond when their pairwise relationship depicted in Fig. 3 is similar w.r.t.: (1) distance $\|\mu_i - \mu_j\| \pm \varepsilon$, (2) angle $\gamma_1 \pm \vartheta$ of \mathbf{n}_i with the line segment joining the two means, (3) angle $\gamma_2 \pm \vartheta$ of \mathbf{n}_j with the line segment joining the two means, (4) angle $\gamma_3 \pm \vartheta$ between both normals projected onto the plane orthogonal to the line segment.

Given two corresponding NDT pairs, we compute two candidate transforms $T_P^Q(i)$ and $T_P^Q(j)$ as follows:

- 1) Compute the rotation R_α that aligns the vectors $\mathbf{v}_1 = (\mu_i - \mu_j)$ and $\mathbf{v}_2 = (\mu_k - \mu_l)$ around the axis $\omega_\alpha = (\mathbf{v}_1 \times \mathbf{v}_2)$ by angle $\alpha = \arccos(\mathbf{v}_1 \cdot \mathbf{v}_2)$.
- 2) Compute the rotation R_β that aligns the normal vectors by rotating along the axis \mathbf{v}_2 . This has two solutions for either NDT pair (i, k) or (j, l) . We obtain $R_{\beta i}$ by projecting $R_\alpha \mathbf{n}_i$ and \mathbf{n}_k onto the plane defined by \mathbf{v}_2 and computing the angle between both projected vectors. We obtain $R_{\beta j}$ analogously using (j, l) .
- 3) Compute the translation \mathbf{t}_i that aligns $\mathbf{v}_1 \times (R_\alpha R_{\beta i})$ with \mathbf{v}_2 . We obtain \mathbf{t}_j analogously for $R_{\beta j}$.
- 4) Return $T_P^Q(i) = (R_\alpha R_{\beta i}, \mathbf{t}_i)$, $T_P^Q(j) = (R_\alpha R_{\beta j}, \mathbf{t}_j)$.

C. NDT Distance Histogram

The naive approach to search for corresponding NDT pairs is to sample randomly from $\text{NDT}(P)$ and $\text{NDT}(Q)$. Winkelbach et al. [34] extend this by placing all evaluated

NDT pairs in a hash map indexed with the four values of their relationship. Thus, on sampling a new pair, they generate transforms for all previously seen ones with the same relationship. We propose to accelerate this further by introducing the NDT distance histogram as shown in Fig. 4. We compute it at the start of the registration process by enumerating and indexing all NDT pairs according to the distances between their means in bins of size ε . Thus, when we search for corresponding NDT pairs, we only sample from those with the same binned distance: $distance-bin(P, d)$ or $distance-bin(Q, d)$. In an idealized setting without sensor noise, scene dynamics, induced error from voxelization, and with 100% scene overlap, the NDT distance histograms for $NDT(P)$ and $NDT(Q)$ are identical. Thus, for each $(NDT_i, NDT_j) \in distance-bin(P, d)$, there are at most $|distance-bin(Q, d)|$ pairs in target to verify and exactly one corresponds to the optimal transform. This property generalizes with some caveats to the practical setting as some NDT pairs do not have any correspondence due to describing non-overlapping parts of the point clouds, or due to imprecise normal computation from sensor noise. We further use the NDT distance histogram to remove pairs with $|distance-bin(Q, d)| = 0$ and to bias the search towards pairs with large distance for more stable alignment, as discussed by Papazov et al. [24] and Aiger et al. [1]. This also exploits the decreasing density of points with distance to the sensor, yielding fewer NDT pair candidates to evaluate as seen in Fig. 4. Our experiments show that convincing results are achieved when we sample from the 25% of bins with largest distance. Thus, in contrast to Winkelbach et al. [34], we front-load the computation of the distances to sample more efficiently and use this knowledge to bias our search towards the most promising pairs.

D. Candidate Transform Evaluation

We designate the application of a transform T to an NDT cell's NDT_i mean and covariance as $T \times NDT_i$. We evaluate how well the candidate transforms align source $NDT(P)$ with target $NDT(Q)$ using the NDT-D2D distance Eq. (6) derived by Stoyanov et al. [29], which we find provides a good trade-off between speed and precision. We index $NDT(P)$ and $NDT(Q)$ with a hash map similarly to Vizzo et al. [32] for efficient correspondence lookup, and denote $h(NDT_i)$ as the cell with the same location as NDT_i in target $NDT(Q)$.

$$\text{dist}(i, j) = -d_1 \exp \left(-\frac{d_2}{2} \mu_{ij}^\top (C_i + C_j)^{-1} \mu_{ij} \right) \quad (6)$$

$$\text{score}(T) = \sum_{\substack{NDT_i \in NDT(P) \\ h(NDT_i) \in NDT(Q)}} \text{dist}(T \times NDT_i, h(T \times NDT_i)) \quad (7)$$

In the remainder, we set the D2D-regularization factors to $d_1 = -1$, $d_2 = 0.05$. We note that $\text{dist}(\cdot) \in [0, 1]$, with perfectly overlapping NDTs have $\text{dist}(\cdot) = 1$ and differences in distance or shape are penalized towards 0. As we iterate over $NDT_i \in NDT(P)$, the score (T) increases monotonously but remains bounded by $N = |NDT(P)|$. We propose to accelerate this evaluation by adapting Capel's [6] bail-out

test to the continuous NDT scores. Intuitively, this test answers the following question: "Given the current score $(T_c)^n$ after n NDTs, will score $(T_c)^N$ surpass score $(T_{\max})^N$, the best transform evaluated so far?" We know from the central limit theorem that the sample mean \bar{x} of i.i.d. random variables with variance σ^2 converges towards the true mean μ with standard deviation $\frac{\sigma}{\sqrt{n}}$, which approaches the normal distribution with larger sample size. We further infer from Popovicius' inequality and $\text{dist}(\cdot) \in [0, 1]$ that $\sigma \leq 0.5$. Thus, after evaluation of n random NDT scores, the 99% confidence bound on the true mean is $\mu = \bar{x} \pm \frac{1.288}{\sqrt{n}}$. We stop scoring the current candidate when it is unlikely to surpass the best transform so far, $\bar{x} + \frac{1.288}{\sqrt{n}} < \mu_{\max}$. As we evaluate transforms, the estimate for μ_{\max} grows, leading to ever earlier bail-out from non promising transforms. For example, we stop evaluating the current hypothesis in our implementation when $\mu_{\max} = 0.9$, $\bar{x} = 0$ and $n = 2$.

E. Exploiting Semantic Information


Many autonomous systems compute the semantic segmentation of incoming point clouds for scene understanding such as the perception systems by Maturana et al. [21] or Hughes et al. [15]. We extend our approach to exploit this additional information for efficiency and to reduce the risk of wrong associations following the ideas of Pfaff et al. [25] and Zaganidis et al. [39]. We split the point cloud according to its semantic classes and compute the NDT and NDT distance map for each class separately, which also effectively reduces the quadratic cost incurred to enumerate all NDT pairs for the distance histogram. We achieve the most convincing results when using NDT pairs from identical semantic classes for both the transform generation and evaluation.

IV. EXPERIMENTAL EVALUATION

The main focus of this work is to present an approach for global point cloud registration. We show the capabilities of our method and support our key claims in our experiments, namely that our approach: (i) performs strongly across different settings; (ii) generates results faster than the state of the art; (iii) optionally leverages semantic information for faster results.


A. Metrics and Baselines


The main metrics to compare global matcher performance are the rotation error (RE) and translation error (TE), calculated as the shortest angular and Euclidean distances between ground truth and the estimated transform. The recall is the proportion of matches where both RE and TE are below a certain threshold. We follow the survey by Huang et al. [14], and set these to the most prevalent values: indoors RE $< 15^\circ$ and TE $< 0.3m$, outdoors RE $< 5^\circ$ and TE $< 2.0m$. We compare our method to the following baselines representing a wide array of state-of-the-art approaches.


TEASER++ : A correspondence-based matcher using FPFH [27] descriptors and TEASER++ by Yang et al. [35]. We use the authors' implementation¹, and set the parameters


¹<https://github.com/MIT-SPARK/TEASER-plusplus>


as Yin et al. [38] outdoors and setting the indoor voxel-size to 0.1 m, normal radius 0.2 m and FPFH radius 0.3 m.



PointDSC  : The state-of-the-art deep-learning based PointDSC correspondence-based matcher by Bai et al. [2] using learned FCGF [9] descriptors. We use the authors' implementation², their trained KITTI-10m model for outdoor settings and their trained 3DMatch model for indoor settings, limiting the number of possible correspondences to 3000.



SC2-PCR  : The state-of-the-art second order spatial compatibility method SC2-PCR by Chen et al. [8] using FPFH [27] descriptors. We use the authors' implementation³, their KITTI-10m parameters for outdoor settings, and their 3DMatch parameters for indoor settings.

IRON  : An NDT feature-based global matcher presented by Schmiedel et al. [28]. We use the authors' implementation and parameters⁴, differing in the following for outdoors (indoors) due to memory constraints from evaluating on large maps: voxel-size 0.5 (0.2) m, matching tolerance $0.5 \times \text{voxel-size}$, neighbor search radius $5.0 \times \text{voxel-size}$, distance/angle bins 5 and RANSAC loops 10000.

Winkelbach  : A dense oriented point matcher proposed by Winkelbach et al. [34]. We use our own implementation and set the following parameters outdoors (indoors), as the authors did not evaluate their approach in these settings: voxel-size 0.4 (0.2) m, normal radius $3 \times \text{voxel-size}$, inlier distance $1.4 \times \text{voxel-size}$, min point pair distance 5.0 (0.5) m, angle hash table bin size: 0.1 rads, hash table bin size distance: voxel-size. We stop after 10s and evaluate the recall of the intermediate results. The main difference to our approach lies in our use of the NDT distance histogram when sampling and NDT-D2D for scoring.

SE-NDT  : A semantically assisted NDT local matcher by Zaganidis et al. [39] with strong global matching performance when used with semantic information. We use our own implementation and set the parameters as they do.

Segregator   : A semantically extended feature-based matcher based on TEASER++ by Yin et al. [38], we use the authors' implementation⁵ and their provided parameters. We evaluate two variants of this semantic approach: one also making use of points labeled as vegetation (Segregator-veg), and the other one ignoring them.

ndt-global (ours)   : Our approach using the following parameters outdoors (indoors): voxel-size 1.0 (0.2) m, $\varepsilon = 0.25 \times \text{voxel-size}$, $\vartheta = 0.1$ rads, min points per NDT voxel 5. We stop after 10s and evaluate the recall of the intermediate results, which shows the duration until the first successful solution and can be used for tuning performance vs time. We distinguish two variants of our approach: one only uses the geometric information (ndt-global), and the other also semantic labels (ndt-global-semantic). On KITTI, ndt-global-semantic only uses points with classes sidewalk, building, fence, vegetation, terrain and pole.

The computation time includes the data preprocessing required to begin the registration process such as feature and NDT computation, but not the semantic segmentation. We run PointDSC, FCGF feature extraction and SC2-PCR on a laptop with an Nvidia Quadro RTX 3000 GPU. All other preprocessing and approaches run on a single thread of an Intel Core i7-10850H @2.70 GHz laptop CPU. This yields a realistic estimate for the compute effort we expect from these approaches in the field, but distorts the computation time in favor of the GPU based methods.

B. Datasets

We consider four datasets and summarize their main characteristics w.r.t. pose perturbation and overlap in Fig. 5. The first benchmark, KITTI-10m, follows Choy et al. [9] and Chen et al. [8]. They use data from runs 8, 9, and 10 from the KITTI odometry dataset [12] recorded with a Velodyne HDL64 and select (scan/pose) pairs separated by at least 10m, yielding 555 registration pairs. We correct for noise in the provided reference poses using ICP from Open3D⁶ as described in the original protocol. For approaches that leverage semantic segmentation, we either generate labels using RangeNet53-512-with-kNN (noisy) by Milioto et al. [23] with a mean Intersection-over-Union value of 41.9%, or use the reference labels (gt) [4]. The benchmark statistics presented in Fig. 5a show that the pose pairs have very low roll and pitch perturbations, and that the challenge mostly stems from the combination of distance/yaw perturbation with overlap that may drop to 40%.

We further use the KITTI-LC 20-30m⁷ dataset to test more complex registration problems in the automated driving context. The dataset statistics in Fig. 5b show that relative to KITTI-10m, the pose pairs are more than twice further apart, and that the scans have lower overlaps. We use the same method to generate semantic labels as for KITTI-10m.

Finally, we also evaluate on the 1400⁸ scan pair global registration benchmark by Fontana et al. [11]. It is a compilation of registration pairs from different datasets, which we further split according to their indoor or outdoor recording setting. This yields 600 indoor instances from the ETH [26] and TUM RGB-D datasets [30], and 800 outdoor pairs from the ETH [26] and Canadian planetary emulation datasets [31]. This benchmark proposes diverse settings from structured to unstructured indoor and outdoor settings recorded with a variety of sensor modalities (LiDAR and RGB-D cameras). The statistics we present in Fig. 5c and Fig. 5d show that the instances in this benchmark are generated with a large variety in roll/pitch/yaw perturbations, while overlap between point clouds is mostly above 40%.

C. Results on KITTI

The first experiment analyzes the point cloud matching capabilities in an automated driving setting and shows that

²<https://github.com/XuyangBai/PointDSC/tree/master>

³<https://github.com/ZhiChen902/SC2-PCR>

⁴<https://github.com/thoschm/IRON>

⁵<https://github.com/Pamphlett/Segregator>

⁶(200 iterations, 10^{-8} relative fitness/rmse)

⁷<https://github.com/HKUST-Aerial-Robotics/LiDAR-Registration-Benchmark>

⁸without KAIST urban 05 due to ground truth issues

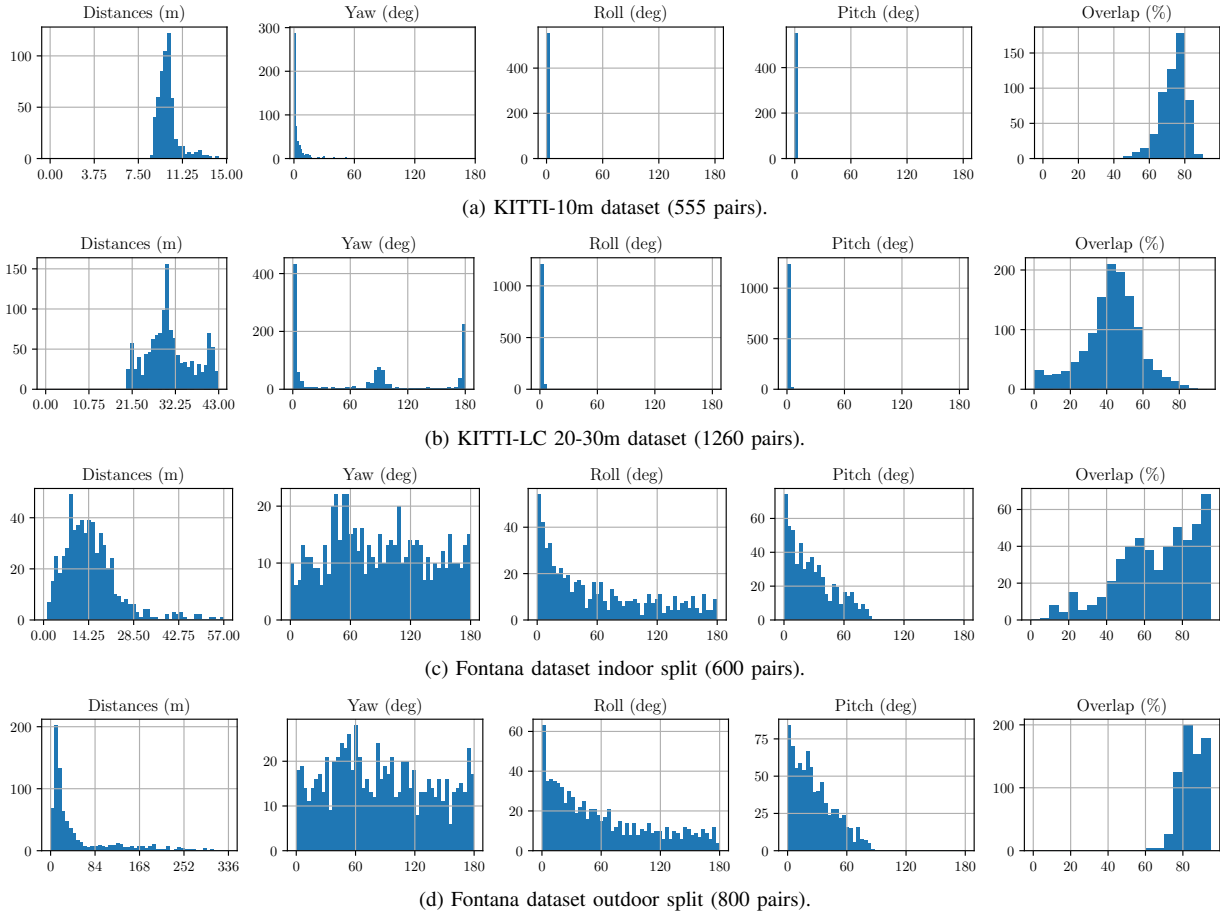


Fig. 5: Overview of the pose perturbations and point cloud overlaps present in the datasets. We compute the overlap as the proportion of points in source having a corresponding point within 0.05 m (indoors) or 0.3 m (outdoors) in target at the ground truth pose.

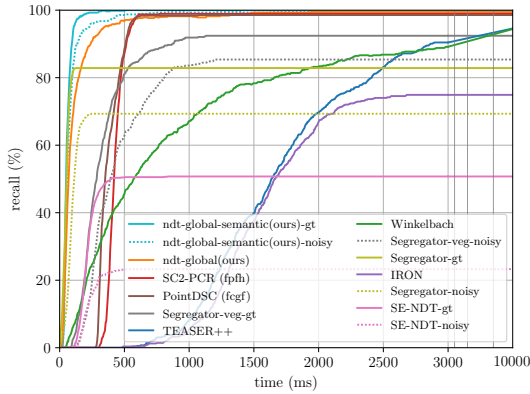


Fig. 6: Registration recall over time for KITTI-10m. Approaches that use semantic information are represented as both solid lines (gt) and dotted lines (noisy labels).

our approach performs strongly in this setting, generates results faster than the state of the art and leverages semantic information for faster results, leading to a support of our three claims. We compare the approaches by plotting the achieved recall over the average registration time required to reach it for the KITTI-10m dataset in Fig. 6. This figure shows that our geometric approach *ndt-global* requires an average of 500 ms per registration pair to achieve 96% recall.

From a convergence speed point of view, this experiment shows that our approach *ndt-global-semantic* leverages pixel-wise semantic information to achieve 99.8% recall in 400 ms, which is faster than any other approach we compare to. While *Segregator* converges at a similar rate, it maxes out at 83% recall. The evaluation also shows that our approach runs faster than the GPU-based baselines *PointDSC* and *SC2-PCR* on this dataset while achieving similar recall values. Our experiments on KITTI-10m Fig. 6 further show our approach to be more resilient against label deterioration than the semantic baselines, dropping from 99.8% to 99.6% recall, while *Segregator* drops from 83.0% to 67.5%, *Segregator-with-veg* drops from 92.7% to 83.8%, and *SE-NDT* from 50.3% to 23.0%. The feature-extracting baselines *TEASER++*, *FGR*, and *IRON* are slow to generate descriptors on large point clouds, explaining why their recall only starts rising after 500 ms, at which point our approach has almost converged to its maximum recall. We observe that *TEASER++* and *Winkelbach* reach recalls $\geq 90\%$, showing the rich geometric information for global registration contained in KITTI-10m.

We present the results of our experiments on the KITTI-LC 20-30m split in Fig. 7, which shows the recalls over the computation time required to achieve them. We first note that all methods have lower recall and a wider spread than on the KITTI-10m dataset, which allows a more differentiated

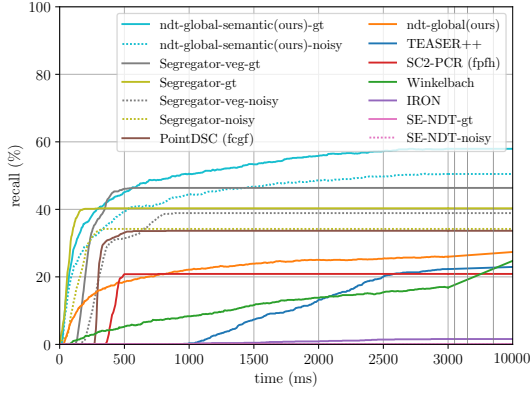


Fig. 7: Results on KITTI-LC 20-30m dataset. Approaches that use semantic information are represented as both solid lines (gt) and dotted lines (noisy labels).

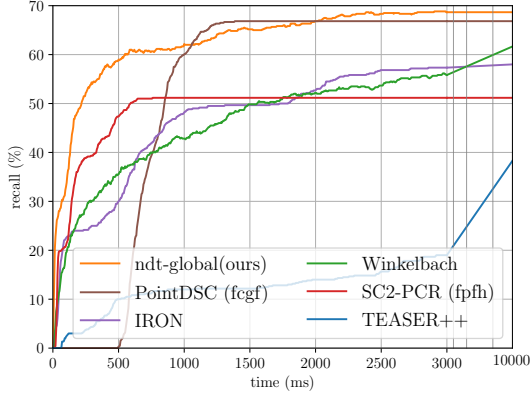


Fig. 8: Results on the indoor split of the Fontana benchmark.

analysis. Interestingly, the semantic based approaches (except SE-NDT) are stronger in this setting than the deep-learned or geometric baselines, which emphasizes the usefulness of semantic information for global point cloud registration as the target scenario grows more complex. Our method, ndt-global-semantic, reaches the highest recall of respectively 58% and 51% using ground-truth or noisy labels. Our method without semantics, ndt-global, also performs better than the other geometric baselines at 28% recall. The factor of two separating the recall of ndt-global from ndt-global-semantic further highlights the usefulness of semantic information.

Summarizing the experiments on the KITTI-10m and KITTI-LC 20-30m dataset, the evaluation suggests that our approach performs strongly in the automated driving point cloud registration scenario, that it generates results faster than the state of the art, and that it also leverages noisy semantic information for obtaining more accurate and faster results.

D. Results on the Fontana benchmark

We present the results of our experiments on the indoor split of the Fontana benchmark in Fig. 8, which shows the recalls over the computation time required to achieve them. Our method, ndt-global, reaches the highest recall of 68%, with 50% reached after 250ms. The only baseline with a similar recall is PointDSC, reaching 67% recall after 900ms processing time on the GPU. Our approach preprocesses

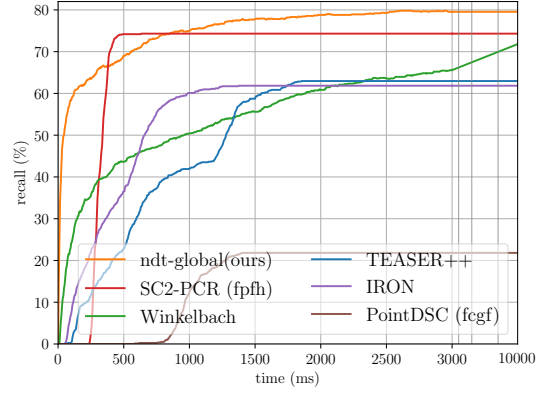


Fig. 9: Results on the outdoor split of the Fontana benchmark.

the data for 22 ms to generate the NDT representation and distance histogram, yielding an average of 480 NDTs per problem. The PPFH based approaches TEASER++ and SC2-PCR have the lowest recall on this benchmark, showing the complexity of tuning this descriptor in diverse scenarios. This experiment also shows that the deep learning approach PointDSC with FCGF features, trained on 3DMatch, transfers well to other indoor scenarios recorded from different modalities.

We present the results of our experiments in the outdoor split of the Fontana benchmark in Fig. 9. Our method reaches the highest recall of 80% after 2500ms, with 50% recall achieved after 100ms. SC2-PCR with GPU computation is the best state-of-the-art baseline reaching 75% recall 500ms faster than our approach. It is particularly interesting to note that the deep learnt model trained on KITTI used in PointDSC does not handle the scale of roll/pitch perturbations present in this dataset, reaching 22% recall. We note that on average, our approach processes the data for 23 ms to generate the NDT representation and distance histogram, yielding 860 NDTs per problem. The PPFH based approach TEASER++ averages 173 ms to compute the descriptors, by this time our approach already has a recall of over 50%.

The experiments on the Fontana dataset suggests that our method provides better registration results substantially faster than most of the state of the art across multiple indoor and outdoor settings with different sensing modalities, supporting our first two claims.

In summary, our evaluation suggests that our method provides fast and robust registration results across all settings and also compares favourably to the baselines. Our evaluation further shows that our approach makes good use of even degraded semantic labels, proving to be more robust to label degradation than the baselines.

V. CONCLUSION

In this paper, we present a novel approach for globally registering point clouds. We build upon the NDT and oriented point pairs framework for candidate transform generation. Our main novelty is the introduction of the NDT distance histogram to focus the search for matching NDT pairs, and our use of pixel-wise semantic information for

greater speed. We implemented and evaluated our approach on different datasets and provided comparisons to other existing techniques and supported all claims made in this paper. The experiments suggest that our approach generates results faster than the state of the art, while effectively using information as gained from semantic segmentation, also when these deteriorate. We believe this work will improve localization and mapping systems by providing compute-efficient, robust and precise global registration.

REFERENCES

- [1] D. Aiger, N. Mitra, and D. Cohen-or. 4-Points Congruent Sets for Robust Surface Registration. *ACM Transactions on Graphics*, 27(3), 2008.
- [2] X. Bai, Z. Luo, L. Zhou, H. Chen, L. Li, Z. Hu, H. Fu, and C. Tai. PointDSC: Robust Point Cloud Registration using Deep Spatial Consistency. *Proc. of the IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2021.
- [3] D. Barath and J. Matas. Graph-Cut RANSAC. In *Proc. of the IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2018.
- [4] 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 Intl. Conf. on Computer Vision (ICCV)*, 2019.
- [5] P. Biber and W. Straßer. The normal distributions transform: A new approach to laser scan matching. In *Proc. of the IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems (IROS)*, 2003.
- [6] D. Capel. An Effective Bail-out Test for RANSAC Consensus Scoring. In *Proc. of British Machine Vision Conference (BMVC)*, 2005.
- [7] X. Chen, A. Milioto, E. Palazzolo, P. Giguère, J. Behley, and C. Stachniss. SuMa++: Efficient LiDAR-based Semantic SLAM. In *Proc. of the IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems (IROS)*, 2019.
- [8] Z. Chen, K. Sun, F. Yang, and W. Tao. SC2-PCR: A Second Order Spatial Compatibility for Efficient and Robust Point Cloud Registration. In *Proc. of the IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2022.
- [9] C. Choy, J. Park, and V. Koltun. Fully convolutional geometric features. In *Proc. of the IEEE/CVF Intl. Conf. on Computer Vision (ICCV)*, 2019.
- [10] A. Drory, R. Giryes, and S. Avidan. Stress-Testing Point Cloud Registration on Automotive LiDAR. In *Proc. of the Advances in Neural Information Processing Systems Workshops*, 2022.
- [11] S. Fontana, D. Cattaneo, A. Ballardini, M. Vaghi, and D. Sorrenti. A Benchmark for Point Cloud Registration Algorithms. *Journal on Robotics and Autonomous Systems (RAS)*, 140:103734, 2021.
- [12] A. Geiger, P. Lenz, and R. Urtasun. Are we ready for Autonomous Driving? The KITTI Vision Benchmark Suite. In *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2012.
- [13] S. Gupta, T. Guadagnino, B. Mersch, I. Vizzo, and C. Stachniss. Effectively Detecting Loop Closures using Point Cloud Density Maps. In *Proc. of the IEEE Intl. Conf. on Robotics & Automation (ICRA)*, 2024.
- [14] X. Huang, G. Mei, J. Zhang, and R. Abbas. A Comprehensive Survey on Point Cloud Registration. *arXiv preprint*, arXiv:2103.02690, 2021.
- [15] N. Hughes, Y. Chang, and L. Carlone. Hydra: A Real-time Spatial Perception System for 3D Scene Graph Construction and Optimization. In *Proc. of Robotics: Science and Systems (RSS)*, 2022.
- [16] H. Lim, B. Kim, D. Kim, E. Lee, and H. Myung. Quatro++: Robust Global Registration Exploiting Ground Segmentation for Loop Closing in LiDAR SLAM. 2023.
- [17] H. Lim, S. Yeon, S. Ryu, Y. Lee, Y. Kim, J. Yun, E. Jung, D. Lee, and H. Myung. A Single Correspondence Is Enough: Robust Global Registration to Avoid Degeneracy in Urban Environments. 2022.
- [18] M. Magnusson. *The Three-Dimensional Normal-Distributions Transform - an Efficient Representation for Registration, Surface Analysis, and Loop Detection*. PhD thesis, Örebro University, 2009.
- [19] M. Magnusson, A. Lilienthal, and T. Duckett. Scan registration for autonomous mining vehicles using 3D-NDT. *Journal of Field Robotics (JFR)*, 24(10):803–827, 2007.
- [20] J. Matas and O. Chum. Randomized RANSAC with Sequential Probability Ratio Test. In *Proc. of the IEEE/CVF Intl. Conf. on Computer Vision (ICCV)*, 2005.
- [21] D. Maturana, P. Chou, M. Uenoyama, and S. Scherer. Real-Time Semantic Mapping for Autonomous Off-Road Navigation. In *Field and Service Robotics*, 2017.
- [22] N. Mellado, D. Aiger, and N. Mitra. Super 4PCS Fast Global Pointcloud Registration via Smart Indexing. In *Computer Graphics Forum*, volume 33, pages 205–215, 2014.
- [23] A. Milioto, I. Vizzo, J. Behley, and C. Stachniss. RangeNet++: Fast and Accurate LiDAR Semantic Segmentation. In *Proc. of the IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems (IROS)*, 2019.
- [24] C. Papazov, S. Haddadin, S. Parusel, K. Krieger, and D. Burschka. Rigid 3d Geometry Matching for Grasping of Known Objects in Cluttered Scenes. *Intl. Journal of Robotics Research (IJRR)*, 31(4):538–553, 2012.
- [25] P. Pfaff, R. Triebel, C. Stachniss, P. Lamon, W. Burgard, and R. Siegwart. Towards Mapping of Cities. In *Proc. of the IEEE Intl. Conf. on Robotics & Automation (ICRA)*, Rome, Italy, 2007.
- [26] F. Pomerleau, M. Liu, F. Colas, and R. Siegwart. Challenging data sets for point cloud registration algorithms. *Intl. Journal of Robotics Research (IJRR)*, 31(14):1705–1711, 2012.
- [27] R. Rusu, N. Blodow, and M. Beetz. Fast point feature histograms (fpfh) for 3d registration. In *Proc. of the IEEE Intl. Conf. on Robotics & Automation (ICRA)*, 2009.
- [28] T. Schmiedel, E. Einhorn, and H. Gross. IRON: A Fast Interest Point Descriptor for Robust NDT-Map Matching and its Application to Robot Localization. In *Proc. of the IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems (IROS)*, 2015.
- [29] T. Stoyanov, M. Magnusson, H. Andreasson, and A.J. Lilienthal. Fast and accurate scan registration through minimization of the distance between compact 3D NDT representations. *Intl. Journal of Robotics Research (IJRR)*, 31(12):1377–1393, 2012.
- [30] J. Sturm, N. Engelhard, F. Endres, W. Burgard, and D. Cremers. A Benchmark for the Evaluation of RGB-D SLAM Systems. In *Proc. of the IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems (IROS)*, 2012.
- [31] C.H. Tong, D. Gingras, K. Larose, T. Barfoot, and E. Dupuis. The Canadian planetary emulation terrain 3D mapping dataset. *Intl. Journal of Robotics Research (IJRR)*, 32(4):389–395, 2013.
- [32] I. Vizzo, T. Guadagnino, B. Mersch, L. Wiesmann, J. Behley, and C. Stachniss. KISS-ICP: In Defense of Point-to-Point ICP – Simple, Accurate, and Robust Registration If Done the Right Way. *IEEE Robotics and Automation Letters (RA-L)*, 8(2):1029–1036, 2023.
- [33] O. Vysotska and C. Stachniss. Lazy Data Association For Image Sequences Matching Under Substantial Appearance Changes. *IEEE Robotics and Automation Letters (RA-L)*, 1(1):213–220, 2016.
- [34] S. Winkelbach, S. Molkenstruck, and F. Wahl. Low-cost Laser Range Scanner and Fast Surface Registration Approach. *Pattern Recognition: 28th DAGM Symposium*, 2006.
- [35] H. Yang, P. Antonante, V. Tzoumas, and L. Carlone. Graduated non-convexity for robust spatial perception: From non-minimal solvers to global outlier rejection. *IEEE Robotics and Automation Letters (RA-L)*, 5(2):1127–1134, 2020.
- [36] H. Yang, J. Shi, and L. Carlone. TEASER: Fast and Certifiable Point Cloud Registration. *IEEE Trans. on Robotics (TRO)*, 37(2):314–333, 2020.
- [37] H. Yin, X. Xu, S. Lu, X. Chen, R. Xiong, S. Shen, C. Stachniss, and Y. Wang. A survey on global lidar localization: Challenges, advances and open problems. In *Intl. Journal of Computer Vision (IJCV)*, 2024.
- [38] P. Yin, S. Yuan, H. Cao, X. Ji, S. Zhang, and L. Xie. Segregator: Global Point Cloud Registration with Semantic and Geometric Cues. *Proc. of the IEEE Intl. Conf. on Robotics & Automation (ICRA)*, 2023.
- [39] A. Zaganidis, L. Sun, T. Duckett, and G. Cielniak. Integrating Deep Semantic Segmentation Into 3-D Point Cloud Registration. *IEEE Robotics and Automation Letters (RA-L)*, 3(4):2942–2949, 2018.
- [40] A. Zeng, S. Song, M. Niessner, M. Fisher, J. Xiao, and T. Funkhouser. 3DMatch: Learning Local Geometric Descriptors from RGB-D Reconstructions. In *Proc. of the IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2017.
- [41] X. Zhang, J. Yang, S. Zhang, and Y. Zhang. 3D Registration with Maximal Cliques. In *Proc. of the IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR)*, June 2023.